WHICH IS MORE?

DEVELOPMENTAL STUDIES ON THE PROCESS OF ESTIMATING RELATIVE QUANTITIES

A. W. SMITSMAN

WHICH IS MORE?

DEVELOPMENTAL STUDIES ON THE PROCESS OF ESTIMATING RELATIVE QUANTITIES

PROMOTORES:

PROF. DR. F. J. MÖNKS
PROF. DR. ANNE D. PICK

WHICH IS MORE?

DEVELOPMENTAL STUDIES ON THE PROCESS OF ESTIMATING RELATIVE QUANTITIES

PROEFSCHRIFT

ter verkrijging van de graad van doctor in de Sociale Wetenschappen aan de Katholieke Universiteit te Nijmegen op gezag van de Rector Magnificus Prof Dr. P. G. A. B. Wijdeveld volgens besluit van het College van Decanen in het openbaar te verdedigen op vrijdag 8 februari 1980 des namiddags te 2.00 uur precies

door

ADRIANUS WILLEM SMITSMAN

geboren te Zwijndrecht

KRIPS REPRO MEPPEL
1980

Aan Coby, Anneloes en Nienke

Aan mijn ouders

ACKNOWLEDGEMENTS

The studies reported in this dissertation were carried out at the Department of Psychology (Developmental Section) at the Catholic University of Nijmegen. I am indebted to several colleagues for their support, advice, comments, and criticism at different stages of this research. Thanks are also due to dr. C.A.J. Vlek for his initial advice, and to drs. H.C. v.d. Hoeven for providing statistical help. I am especially indebted to prof.dr. E.J. Mason for his generous assistance in the preparation of this manuscript. I would also like to thank the several students without whose help and cooperation the experiments could not have been run. I would like to thank particularly drs. J.M. Sterenborg who as a graduate student prepared most of the computer programs for running the experiments. I would also like to thank the Computer Section of the Psychological Department for their technical support. Appreciation is also expressed to parents, children and teachers of the nursery schools 'Maria Theresia', 'Hummeltjeshonk', and ''tKraayenest', and the elementary school 'Henri Dunant', all in Nijmegen or the neighbourhood of Nijmegen. I am grateful to Lenie Vos, Ricky v.d. Heyden, and Lea Wagner for typing earlier versions of this manuscript, and to Han van Valderen for typing the final version. I would also like to thank Ilja ten Brink at the Audiovisual Service of the Faculty of Social Sciences, for drawing the figures. Above all I thank my wife Coby for her assistance and interest and my little daughters Anneloes and Nienke, who have been an incessant source of inspiration.

CONTENTS

1. GENERAL BACKGROUND AND PURPOSE OF THE STUDIES

1.1. Introduction

Children and adults daily use terms like 'more', 'less', 'much', and 'often' to label their estimations of quantity. On the basis of such estimations inferences may be made about objects, events or outcomes. For example, Estes (1976) showed that predictions of future outcomes were based on the estimated relative frequency of past events, e.g. wins and losses by candidates in an election.

Generally, a quantitative judgment may refer to continuous (e.g. volume, mass, weight) or discontinuous (e.g. collections of objects) quantities. Furthermore, a continuous quantity can be made discontinuous by employing some measurement procedure, that is, by dividing it into discrete sections or units, according to some rule. The present studies were concerned with the comparison of discontinuous quantities or sets of elements. Therefore, in these studies the term quantity refers to those types of quantities.

A quantitative comparison like 'more' is a very simple but basic relative quantity judgment. A relative quantity judgment requires at least some representation of the quantities involved, and a set of rules in order to determine their relative magnitude. In case of a quantitative comparison a relation of inequality, or more specifically, a relation of more, less, or equal, has to be established (see e.g. Klahr & Wallace, 1976).

In order to represent quantity number may be considered as very precise and useful providing the opportunity for exactness. However, other and less precise forms of representation may also be used by people. For example, young children ordinarily will base a quantitative comparison of simultaneously presented collections of objects on spatial representations or cues. In line with this finding, it may be argued that the exactness of number is guaranteed only if, for example, all elements of the collections involved are counted correctly. At an early age, children may perhaps understand the basic principles of counting. Nevertheless, they make a lot of mistakes in applying those principles (see e.g. Gelman, 1972; Gelman & Gallistel, 1978). Therefore, probably number is not used by children until exact quantification procedures can be applied skilfully.

Aside from skill, situations may make it also impracticable to count all elements and consequently to get an exact numerical representation of each of the collections involved (see e.g. Beckwith & Restle, 1967, and Gelman & Gallistel, 1978 for a definition of counting). For example, the number of elements may be too large and/or the exposure time too short. Moreover, a subject may be distracted and consequently inspect a collection too briefly. In ordinary life, numerous factors may prevent a person from forming exact numerical representations. In those situations a person has to estimate. At present, this process of estimation is still puzzling. That is, it is not clear what types of representations are formed in those situations and how these representations may be acquired.

According to Klahr & Wallace (1976), idiosyncratic representations are formed referring to spatial characteristics in case of spatially presented quantities. According to them, for example: "some people estimate length in terms of football fields, others in terms of cars" (p. 65). From a developmental point of view it may be argued that number may also be used when estimation is required by the quantification situation. As children grow older they learn about quantities and quantification. For example, at some age, it is understood how the numerical value of a collection of objects can be changed and how not. Moreover, the usefulness of number for representing and comparing quantities will be comprehended. That is, when number becomes of central significance to quantitative thinking by individuals it may be expected that reliance on number will not be restricted to situations that permit exact quantification.

1.2. General background

Estimation of discontinuous quantities or sets of elements has been studied within psychology from several different theoretical perspectives, among them psychophysics, statistical decision-theory and verbal learning.

First, within psychophysics studies were directed at the perception of numbers. A distinction was made between numerousness and numerosity. Numerousness was defined as a physical property of a collection of objects which can be perceived without counting, that is by estimation. It was contrasted to numerosity, the number that can be determined, for example, by counting (see e.g. Kaufman, Lord, Reese & Volkmann, 1949). The distinction between numerousness and numerosity roughly corresponds with that between relatively large and relatively small numbers. In order to investigate numerousness perception, random patterns were presented of varying numbers of dots. Numbers were at least of such a size that counting of all the elements within a pattern could not be done within a given exposure time. Although, generally, subjects increasingly underestimated the number presented, results suggested that, on the average, people are fairly accurate in estimating the number of dots sometimes varying from 12-1500 dots (see e.g. Guttman, 1978; Krueger, 1972; Wing, 1971).

Second, according to statistical decision-theory, rational predictions, inferences, expectations, etc., are completely or partly based on observed frequencies and relative frequencies (see e.g. De Finetti, 1974[a], 1974[b]; Peterson & Beach, 1967; Vlek & Wagenaar, 1976). For example, the relative frequency of accidents or near accidents in nuclear plants during the last decennia is considered as a relevant source of information in order to judge the risk of civil application of nuclear power. From the perspective of statistical decision-theory, studies were concerned with the ability of individuals to estimate relative frequencies or frequency distributions. A conclusion seems to be that the relative frequency of past events, not only in binary, but also in multiple probability learning tasks of up to ten different categories, can be estimated fairly accurately (see e.g. Vlek, 1970; Vlek & Werner, 1973).

Finally, within the context of verbal learning, studies were concerned with aspects of memory storage of repetition or successive presentations of the same items. It is suggested that information about the frequency of past occurrences or co-occurrences of objects and events may be available from memory (Anderson & Bower, 1972; Hintzmann & Block, 1971; Underwood, 1966), and may be used, for example, in discriminating and recognizing objects and events (Ekstrand, Wallace & Underwood, 1966). It has been demonstrated that the frequency of a word contained in a list, separately or as a paired associate, can be reproduced reasonably well by human subjects, whether or not they knew at the time of presentation that frequency would be tested (e.g. Howell, 1973).

Although the number of studies on relative quantity estimation is steadily increasing, especially within the context of verbal learning, how sets of elements are estimated by human beings, and the conditions that may affect the accuracy of them is not yet clear. In most estimation studies items were presented sequentially. However, investigation of the estimation process in sequential tasks seems to be rather complicated, because the way estimation may be described psychologically depends on the way repetition is stored in memory.

Two major hypotheses about representation of repetition are: (1) the multiple trace hypothesis and (2) the frequency-attribute hypothesis (see e.g. Howell, 1973 for an overview). The first hypothesis holds that repeated presentations or repetitions of the same items will be stored in memory in the form of separate traces (Hintzmann & Block, 1971) or 'list markers' (Anderson & Bower, 1972). According to the second hypothesis, however, frequency is considered to be an attribute of memory (Underwood, 1969). That is, repetition of the same item is stored in a form analogous to a counter. Within the context of the first hypothesis, it is suggested that frequencies will be quantified by counting traces or 'list markers' for small frequencies, and by some sampling and estimation procedure for larger ones. Within the context of the second hypothesis, various tentative suggestions are made about how the counter may operate. For example, Begg & Rowe (1972) hypothesized a kind of paired-associate learning in which after a repetition of an item its associate, a number representing the frequency of earlier repetitions, will be incremented by one. Another suggestion was that the counter is a kind of automatic processing, or cataloguing operation (Hasher & Chromiak, 1977). Finally, according to Begg (1974) an internal representation of frequency, the frequency counter, may be inferred by counting, sampling and/or estimating items of the same type at the moment of *presentation*. So, the quantification process suggested by Begg may be comparable to the one suggested by the multiple trace hypothesis, except for the moment of quantification. According to the first hypothesis quantification is done during presentation, but according to the latter one, during retrieval of individual items. In both cases, however, *collections* must be quantified.

A heuristic way of estimating frequencies and relative frequencies, in line with the multiple trace hypothesis, is suggested by Tversky & Kahneman (1973). These authors suggested that frequencies and relative frequencies may be estimated by assessing the availability of the relevant category instances; that is, "by assessing the ease with which the relevant mental operation of retrieval, construction, or association can be carried out" (p. 208). Application of this method of estimation, which they called the availability heuristic, may lead to systematic biases. For example, in one study subjects were asked to estimate the relative frequency of English words with a given letter in the first versus in the third position. For example, the frequency of words of at least three letters beginning with K was estimated to be much higher than that of words with K in the third position, although there are twice as many words with K in the third as with K in the first position. Presumably the reason for the bias is that subjects can more easily recall instances of words beginning with K than of words with K in the third position.

In a series of 10 experiments Tversky & Kahneman demonstrated that the availability of individual items may affect the estimation of their frequency or relative frequency. That is, the relatively more available items were overestimated. However, it was not made clear in these studies whether availability was assessed in reality by the subject, that is, whether the estimation was based on perceived availability. In the tasks presented, individual instances were either not shown, but could be constructed by the subject using a rule suggested by the experimenter or were presented sequentially.

According to Tversky & Kahneman availability may be assessed by constructing (the first type of task), or by retrieving some instances for a short period of time (the second type). However, when individual items are constructed or retrieved an estimation may be based on the composition of the acquired sample. A possible explanation for an availability effect may be made in terms of the construction rule or retrieval process used, which sometimes generates non-representative samples. Besides, it is also suggested that availability may be assessed directly, that is without generating instances. However, that this would be accomplished is not clear. Finally, there are no compelling rational arguments for the assessment of availability, that is, using a rule that attributes a higher frequency to categories with relatively greater availability. Yet, many factors unrelated to frequency may effect availability, e.g. recency. Furthermore, Beyth-Marom & Fischhoff (1977) showed that some categories whose instances appeared to be unavailable to the subject, were estimated to be relatively higher in frequency of occurrence.

In the foregoing paragraphs the discussion of psychological investigation of estimation was based predominantly on studies using adults as subjects. In regard to children, Fischbein (1975) has stated that a 'relative frequency intuition' that regulated the storage and use of frequency information is available to children as early as the age of three or four years. This claim was based on results of probability learning studies with children (see also Goolet & Goodwin, 1970, for an overview). Frequency estimation is also supposed by Piaget & Inhelder (1951) in the context of the development of the concept of probability. However, the way frequencies and relative frequencies are estimated and the development of the estimation process is beyond the scope of Fischbein's and Piaget & Inhelder's studies. Both were directed to the mathematical aspects of coping with frequency information. Furthermore, processes that yield a quantitative representation, or quantification mechanisms, are generally not studied within the context of Piaget's theory (see Flavell, 1977; Gelman, 1972, Klahr, 1973; Klahr & Wallace, 1976). It has been suggested by Macnamara (1975) that the quantification mechanism known as counting was not studied within this context because of the Piagetian concern to reduce number to logic.

In summary, the literature does not clearly describe the estimation processes. The situation is even less clear regarding the role of development or growth of cognitive mechanisms in children. The present studies were designed to explore the processes involved in estimating relative quantity, and the differences in these processes at various ages.

1.3. General purpose of the present studies

The present studies were designed to investigate:

1) how the relative quantity of simultaneously presented collections is estimated by persons of different ages;
2) whether the estimation process does change with age, that is, within the range from about six years to adulthood.

More specifically it was examined whether number will be used in quantitative comparisons of collections when estimation is required, and whether the use of number will depend on age.

In order to investigate these questions a simultaneous method of presenting collections was selected for the present research instead of a sequential one. Although local effects may occur with both methods of presentation, simultaneous presentation offers more experimental control over the quantity upon which an estimation is based than sequential presentation. When items are presented sequentially estimations may be inferred from currently presented items, others from the past, or a combination, and it might differ for persons of different ages. Moreover, children of different ages may remember more or less of sequentially presented items, and consequently age differences in estimation and in memory may be confounded.

The task used in the present research was designed to represent a simple relative quantity estimation task, determining which of two categories, circles or squares, occurs most frequently in an array. With this type of task, understanding of the concept 'more' by subjects at each age level is relatively easy to determine, and provides a sufficient basis for the interpretation of estimations when compared, for example, to understanding of numbers or proportions.

The following chapters describe the present work in detail. First, chapter 2 discusses general and developmental studies relevant for the processes of estimation of relative quantities in arrays of objects or figures, and proposes a process of estimation that is based on number. In chapter 3 the first of four studies which investigated the proposed process in children and adults is presented. Chapter 4 describes three follow up studies investigating different aspects of the proposed process and its application at different ages in more detail. Chapter 5 discusses the main results of the four studies.

2. THE ESTIMATION OF SIMULTANEOUSLY PRESENTED RELATIVE QUANTITIES: AN INTRODUCTION

Since its beginning experimental psychology has dealt with the perception of numbers. One of the earliest studies in this area was by Jevons (1871). Since the Jevons research, studies in this area have appeared periodically, although the number of studies accumulated in the literature is not large. Recently, there seems to be renewed interest in this field, especially in the perception of small numbers by adults and children (see e.g. Atkinson, Campbell & Francis 1976[a], 1976[b]; Chi & Klahr, 1975; Klahr & Wallace, 1976; Svenson & Sjöberg, 1978).

Studies of the estimation of relative quantities of objects have been even fewer than those of the perception of numbers. The majority of these studies used adult subjects and investigated proportion estimation. The scarcity of research with children as subjects seems rather remarkable because of the considerable interest in the study of quantitative comparison among developmental psychologists. Furthermore, studies directed at the process of estimation of relative quantities are generally not available in the literature.

In this chapter, what is known about the process of relative quantity estimation and its development is explored. The chapter is divided into four sections. In the first, relevant studies on quantitive comparison in children will be discussed in order to investigate how children may estimate and what type of changes in estimation may be expected with age. The second section is directed toward proportion estimation studies in adults. Although in proportion estimation the comparative judgment is relatively more complex than in quantitative comparison, quantitative representations are perhaps formed in similar ways. The sampling hypothesis derived from studies on proportion estimation may be of relevance in this respect. This hypothesis will be discussed, especially with regard to the units of sampling. The third section explores the possibility that the sampling units consist of small groups of objects. This discussion is based upon evidence presented in the first two sections, and from studies on the perception of small numbers. Finally, in the last section, research questions will be stated and the general design of the studies reported in the next chapters will be given.

2.1. Quantitative comparison and estimation in children

Quantitative comparison, or the ability to compare at least two collections in some quantitative sense, seems to be one of the first number skills human beings develop (e.g. Siegel, 1971; Schaeffer, Eggleston & Scott, 1974). Children of about three years seem to be able to judge which collection contains more objects or is larger. It can be easily observed by giving one child more candies or toys than a friend or sibling.

Within developmental psychology, many studies have been directed at quantitative comparison. Results of studies involving collections will be considered here in relation to estimation, especially the developmental aspects of this process.

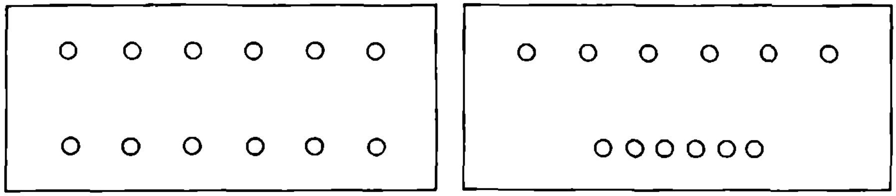
In quantitative comparison studies, ordinarily, linear arrays of objects were presented. Linear and random arrangements of nine or less dots were used by Siegel (1972) in a discrimination learning study. She showed that children ranging in age from 3-0 to 4-11 could be taught to discriminate between collections differing in number of objects. Furthermore, random arrangements appeared to be more difficult than linear ones, especially for the younger children.

In Siegel's study and in nearly all relevant studies, the number of objects presented was very small, usually ranging from two to ten, and estimation was not definitely a requirement unless, for example, the counting skill of a child was insufficient with respect to the given numbers. However, estimation was required in a study of Hecox and Hagan (1971), using random arrangements of 100 figures of two types. In this study the accuracy of proportion estimation was studied in children from five to seven years. It was found that even the youngest children could discriminate between different proportions, ranging form .10 to .90. Moreover, accuracy of discrimination increased with age. However, investigation of the way in which comparisons were made was neither a part of this nor of Siegel's study.

Quantitative comparison has been examined from different perspectives. The simplest approach was to investigate how large a collection children of various ages could quantify correctly by specifying the number of it or by determining the largest or smallest one of two or more collections presented. An extensive review of this type of study has been given by Gelman (1972), but will not be discussed here since the results do not specifically clarify the process of estimation used by children. Studies of conservation of number and studies derived from this paradigm seem of more relevance to the present investigation.

In the Piagetian paradigm for investigating number conservation, two linear arrays of objects are shown to the child. The arrays are arranged so that the numerical equivalence of both collections can be determined easily by one-to-one comparison (see Figure 2.1[A]). After the numerical equivalence is confirmed by the subject, one array is lengthened or shortened by moving the objects in it closer together or further apart (see Figure 2.1[B]). The subject observes this spatial transformation and then is again asked questions concerning the numerical equivalence of the arrays. Subjects who answer these questions correctly are ordinarily called 'conservers' while those who answer that the numbers of objects is different in the new arrangement are called 'non-conserver's. According to Piaget (1952) a child's understanding of the principle of cardinal equivalence of sets can be measured by the conservation of number test (see also Flavell 1963, p. 309-316).

Number conservation appears to develop in children between the ages of five and seven. It has been investigated in many different studies. Reviews of this work have been done recently by Brainerd (1978), and Flavell (1977). Of special relevance to the present subject of investigation are studies concerned with the type of information that is used by non-conservers to make quantitative comparisons. Although estimation was



A B

Figure 2.1. Schematic representation of number conservation before (A), and after (B) a spatial transformation of the bottom row.

not strictly a requirement in those studies, some results are of importance since comparisons made by these children may be achieved really by estimating. Furthermore, a change in the way of estimation, when number conservation is developed in children, is suggested by current conceptualizations (e.g. Gelman, 1969) of the acquisition of number conservation.

Several studies investigated the type of information that is used by non-conservers. The ages of the children in the different studies varied from about three to seven years. Tasks required making quantitative comparisons about untransformed rows of varying length, density, and number. Since length, density and number cannot be varied independently of each other, several triads of rows were constructed and presented by Gelman (1972) and Smither, Smiley & Rees (1974). Pairs of rows, on the other hand, were used by La Pointe & O'Donnell (1974), Lawson, Baron & Siegel (1974), Pufall & Shaw (1972) and Pufall, Shaw & Syrdal-Lasky (1973). In these studies the number of figures within a row was relatively small, for example from two to nine in Gelman's study. Although the different studies are not directly comparable, it may be concluded that length is a more important cue for quantity in young children than density. That is, the number of objects of the longer row was generally estimated to be larger than that of another shorter one, regardless of whether the first contained more, less, or as many objects as the latter. However, comparisons of rows with three or less objects were mostly based on number, that is, made correctly. Besides the tendency to base comparisons on number with rows containing four of more objects increased with age. This can be explained by a change in method of quantification, that is, by a transition of estimation to non-estimation, or more precisely, counting. In fact, in all the studies of this type, rows were presented without any restriction in time, so that comparisons of the relatively few objects could be made by means of counting. According to Gelman (1972), and Gelman & Tucker (1975) overt counting in this type of task decreased with age. However, older children are more skilled in counting than younger ones (Potter & Levy, 1968; Schaeffer et al, 1974). Therefore, when counting is not overt it does not follow that there is no counting at all, but rather that counting is not observed.

In the above studies, a change in the manner of estimating could not be shown, since estimation may not have been required at older ages. One might ask whether the use of spatial information decreases with age using collections too numerous to count. Further, will conservers estimate on the basis of spatial information or some other one? These types of studies are not generally available in the literature. Therefore, suggestions about what may be expected cannot be derived directly from existing data. However, some inferences may be made from conceptualizations of the acquisition of number conservation.

It has been hypothesized that children who fail to conserve do not discriminate between relevant and irrelevant features of quantity (Bryant, 1974; Gelman, 1969; Wallach, Wall & Anderson, 1967; Zimiles, 1963, 1966). Gelman has shown that acquisition of conservation can be considered to be a problem of ignoring irrelevant features (e.g. length), and attending to the relevant feature of quantity (e.g. number). In her study, children ranging in age from 4-6 to six years, were given discrimination training. Specifically different triads of rows, varying in length, density, and number (three to six objects) were presented to children who failed to solve conservation problems. Spatial characteristics were made irrelevant and number was made relevant for discriminating correctly. Nearly all trained children showed complete understanding of number conservation, not only immediately after training, but also 2-3 weeks later. Bryant (1974) has also shown that young children can be easily trained to ignore spatial characteristics of rows.

One may expect that ignoring spatial characteristics affects the way of estimating. That is, children who ignore these characteristics estimate differently than children who consider these relevant. However, there are some interpretation problems with respect to the assertion that children who conserve have learned to consider spatial characteristics as irrelevant to quantity.

First, it is not clear whether spatial characteristics would be considered irrelevant to quantity independent of the type of situation. That is, one might ask whether spatial characteristics would be ignored when there is no need to estimate or when estimation is the only possible way to make comparisons. In the current quantitative comparison literature with children as subjects, there are no studies that compare these two alternatives. Second, one may argue, that understanding of the principle that a change of length is compensated by an opposite change in density in a spatial transformation is essential for understanding number conservation (see e.g. Piaget, 1952; Brainerd, 1978). From this, one might conclude that conservers consider both length and density as relevant to quantity. However, a possible awareness of the fact that a change in length is compensated by an opposite change in density in conservers does not logically require that both length and density be considered as relevant to quantity. On the contrary, both features may be considered completely irrelevant, and number may be used.

Summarizing, children of 5-7 years old already know much about numbers and have developed some number skills (see e.g. Flavell, 1977). They know, for example, the principle of cardinal equivalence (Flavell, 1963) and are able to count up to about 10 or more (see e.g. Schaeffer et al, 1974). In addition numbers of three or four objects can be perceived very quickly. As the method of estimation is developed, it may be expected that existing number knowledge and skill is incorporated in the new strategies of estimation. It may be that these new strategies take number as input.

2.2. Proportion estimation in adults and the sampling hypothesis

As has been stated in the introduction of this chapter, most of the relative quantity estimation studies with adult subjects required estimation of proportion. Therefore, a short description of the way in which proportion estimation has been studied is provided here. Furthermore, the major findings of these studies will be discussed. Finally, the sampling hypothesis in estimation will be reviewed, and three general hypotheses about sampling units will be explored.

Proportion estimation is a more complicated type of judgment than the simple quantitative comparison process just discussed. In proportion estimation studies, different proportions of randomly arranged figures of two types, mostly dots of two colors, are presented. A subject has to judge the proportion of the figures that are of one type. Proportions varied from about .10 to .90 and were presented with short presentation times, varying from 1 second to 10 seconds. An estimation was given in terms of a percentage or on a 7 or 11 point scale.

Two systematic kinds of error tendencies have been found in such studies. Over estimation of low, and under estimation of high proportions was found by Stevens & Galanter (1957). The opposite error tendency, under estimation of low and over estimation of high proportions was found by Shuford (1961). Brooke & MacRae (1977) calculated systematic error tendencies on individual data. Most subjects showed under estimation of low, and over estimation of high proportions, but the opposite tendency, or no systematic error tendency was also found among some subjects. These divergent results have been difficult to explain (see e.g. Peterson & Beach, 1967). Possibly, they are specific to proportion estimation. Different factors may contribute to systematic error tendencies such as the way a relative amount is determined and the response method used (see e.g. Nash, 1964; Pitz, 1965; Wagenaar, 1975). Brooke & MacRae argued that systematic error tendencies are not a function of the response method used. However, the results of their study were not very convincing in this respect. Predictions derived from their hypotheses were supported in only three of the seven cases.

Peterson & Beach (1967) considered proportion estimation as a process of statistical description. They argued that sample information is the only possible form of information that can be extracted from a display of figures. In reviewing sequential as well as simultaneous proportion estimation studies, they concluded that the results of these studies supported their hypothesis. It was found that the accuracy of estimation generally increased with longer presentation times and with more extreme proportion values.

When proportion estimation is considered to be a statistical description task simple sampling that is also unbiased is assumed. In an estimation study of Shuford (1961) proportions, varying from .10 to .90, were represented using stimulus configurations with red and blue squares, and configurations with vertical and horizontal bars. Results suggested that proportions of red figures were over estimated when compared with the same proportions for blue figures. The same effect was found for horizontal, compared with vertical bars. Over estimation was more pronounced for proportions in the midst of the scale than for the more extreme values and decreased with presentation time from 1 to 10 seconds. Shuford explained these results in terms of biased sampling, assuming a statistical or stochastic mechanism in proportion estimation behavior of people. In other situations, it has been demonstrated several times, that this assumption is questionable (see e.g. Kahneman & Tversky, 1972; Tversky and Kahneman, 1973; Wagenaar, 1972; De Zeeuw & Wagenaar, 1974). Furthermore, the term 'biased' explains nothing. That is, it does not explain why elements of a category are over-represented in a sample. It only describes the fact that they are. For explanation, sampling should be defined psychologically. That is, the quantifiable units that are drawn out of a collection, the quantification process given these units, and the inference of a quantity judgment given one or more samples from the collection should be described as psychological processes.

Different quantifiable units may be used by a subject. Theoretically, they might refer to one element (a circle, a bar, etc.) or a group of elements. Moreover, physical features such as visual area or density or both in combination may function as dimensions. One may estimate relative quantities, for example, on the basis of weighted density differences in different parts of a field.

It can be stated that there is more evidence contrary to than favouring the hypothesis that adults estimate relative quantities on the basis of spatial characteristics. Buckley & Gilman (1974) found no evidence that quantitative comparisons of pairs of dot patterns were based on spatial or other physical characteristics. All combinations of two numbers in the range from 1-9 were presented in the form of digits or dot patterns. Subjects judged which of a pair was numerically larger. Analysis of latencies revealed no essential differences for digits or dot patterns, especially for combinations of the numbers four to nine. The stimuli were presented without any time restrictions for responding. Therefore, this study may not have really investigated estimation. In another study, Taves (1946) found differences in estimations of the same random patterns of 14 to 180 dots, on the basis of visual area and number. Krueger (1972) varied the density of random dot patterns for different numbers of dots and concluded that density does not play a central role in number estimation. Features that were related to spatial characteristics were varied by Bevan & Turner (1964), Granberg & Aboud (1969) and Granberg (1972). The results of these studies are difficult to interpret and rather inconclusive.

Dixon (1978) showed that the time a subject requires to make a numerical comparison of pairs of rows containing a different or the same number of figures was affected by some differences in length. The number of figures within a row was very small, two to five. For each pair of numbers presented, the difference in length of both rows was varied. For example, for the pair 2-3, the length of the line having two objects, might be shorter, the same, or longer than that of the three object line. That is, the rows were made incorrespondent by arranging the numerically smaller row to be the relatively longer. For most number combinations latencies increased with the degree of incorrespondence. Dixon explained these results using a complex quantitative comparison model which took brightness, length, and density difference as input. It is highly unlikely however, that the numbers in this study were compared by weighting these dimensions. In the next paragraph it will be shown that these numbers can be perceived correctly and quickly. In Dixon's study, these comparisons were made correctly. Therefore, a more plausible hypothesis for the Dixon data, would be that length interfered with a number comparison process.

The balance of evidence seems to weigh against, rather than in favour of, the significance of spatial characteristics in estimating relative quantities. For the hypothesis that the units of sampling are single elements, sampling is in essence counting or enumerating single elements. It is not probable that relative quantities are estimated by counting because it has been shown that highly accurate proportion estimations can be given in some situations with presentation times of about one second, and such short duration would preclude accurate counting (e.g. Shuford, 1961).

So far there seems to be little definite support for the alternatives to the hypothesis that the sampling units are formed by quantifiable groups of elements. There is on the other hand some positive evidence in favour of this hypothesis. Glanville & Dallenbach (1928), for example, reported that six or more unordered patterns of geometrical figures were estimated by forming and adding groups. Each pattern was presented for 150 msec. and represented a different number of elements ranging from 3 to 15. According to subjects's reports, groups were formed on the basis of an after image of each pattern, and the number of figures within a group was perceived immediately on presentation.

2.3. The sampling of quantifiable groups

At the end of the first paragraph of this chapter, it was proposed that the manner of estimating could change in children between the ages of about five to seven. It was suggested that children at that age may have learned to consider spatial features as irrelevant to number. Therefore, for that age an estimation procedure would be required that ignores spatial information. Further, children of that age know something about numbers and procedures for determining number exactly. Therefore, the new estimation procedure would be developed using the number knowledge and number skills that are available. Thus, number would be considered part of the input.

At the end of the second paragraph of this chapter, it was argued that adults used sampling as the method of estimating by considering groups of objects as quantifiable sampling units. In this section more empirical arguments for such a process will be given. First, the perception of number for small groups of objects will be reviewed. Second, empirical evidence will be considered regarding the hypothesis that relative quantity estimation is based on sampling small quantifiable groups. Third, the general design of the studies of that hypothesis will be explored.

It has been demonstrated several times that up to about five objects can be perceived quickly and correctly, even with a presentation time of 150 msec. Kaufman, Lord, Reese, & Volkmann (1949) introduced the term subitizing to distinguish this process from counting and estimating. Counting was described by them as enumerating (see also Beckwith & Restle, 1966). Estimating, on the other hand, was defined as perception of quantity without counting for seven or more objects. According to these authors, subitizing was possible for up to six objects. Reviews of subitizing studies have been done by Atkinson, Campbell & Francis (1976[a]), Klahr (1973), and Klahr & Wallace (1976). According to Atkinson et al and Klahr subitizing is possible for up to four objects, and according to Klahr & Wallace for up to three objects. The number of objects that can be quantified by subitizing, or the subitizing range, depends on the way in which it is operationalized. That is, speed, accuracy, confidence of judgment, or combinations of these were used to operationalize subitizing in different studies. Moreover, the criteria for speed, accuracy, or confidence differed among the studies. Finally, the content of the subitizing process itself has not been identified. Neisser (1966) suggested some form of pattern perception or automatized counting (see e.g. Vos, 1977). Since subitizing is defined operationally, differences in operationalization have limited the usefulness of attempts to determine the subitizing range or to compare age groups on their subitizing range as was done by Chi & Klahr (1975) and Svenson & Sjöberg (1978). The first authors showed subitizing in children of about five years old and the second in children of about seven to eight years.

Subitizing appears to be a very fast process. The speed of subitizing was calculated by Klahr & Wallace (1976) at about 0.05 sec. per object for adults, at about 0.19 sec. per object for children at the age of five years, and by Svenson & Sjöberg (1978) at about 0.11 sec. for children of 7-8 years. In these studies comparable procedures were used for calculating the subitizing speeds with a range of one to three objects.

Although subitizing has not been defined very well, the general conclusion that may be drawn from studies on subitizing is that small numbers of objects can be quantified quickly and correctly despite very short presentation times. This finding has been obtained with children of about five years old or younger as well as adults (see e.g. Flavell, 1977; Schaeffer et al, 1974). From this and other evidence it may be hypothesized that the quantified units in sampling consist of subitizable groups. Moreover, if sampling is based on subitizing, then it may be hypothesized that estimations can be influenced by the ease with which small groups can be formed from an array of objects.

In experiments on estimation of relative quantities usually groups are not given, but must be discovered and abstracted from larger and randomly arranged collections of objects. Studies of the effect that ease of subitizing has on estimation are generally not available in the literature. However, some results derived from studies of estimation of numbers, enumeration, subitizing, and relative quantity estimation may be interpreted as supporting the possible relevance of subitizing for estimation. In the following paragraph, some of this literature is discussed.

Granberg & Aboud (1969), and later Granberg (1972), reported that accuracy of number estimations decreased when a collection of objects whose number had to be estimated was mixed with a larger collection of objects of another type. It is possible that the formation of groups of relevant objects is hindered by the presence of irrelevant objects. Klahr (1973), and Klahr & Wallace (1976) suggested that in number estimation some starting-value is calculated by forming groups, subitizing, and adding groups in a part of an arrangement. An estimation of the whole can then be derived from the numerical representation of the part. When the presence of irrelevant objects results in a lower starting value for a part of a collection, it may be expected that the estimation for the whole collection will be lower as well.

Atkinson, Campbell & Francis (1976[b]) showed that the number of objects judged correctly at a presentation of 150 msec., could be increased by arranging the objects in easily discriminable groups. In another study by Atkinson et al (1976[a]), it was found that four or less objects were correctly quantified or subitized. In that study groups were not given. In the Atkinson et al (1976[b]) study, eight or less objects were correctly quantified using groups of four or less objects.

Beckwith & Restle (1966) reported that latencies for enumerating 12, 15, 16 or 18 objects were shorter when the objects could be easily divided into groups. According to the authors, number judgments were constructed by adding subitizing groups.

Finally, a perceptual illusion discovered by Frith & Frith (1972) using a solitaire marble board and described as 'the solitaire illusion' can be explained in terms of sampling on the basis of subitizing (see Figure 2.2[A]). In the first pattern in Figure 2.2 it seems as if there are more solid black circles than open ones, although the number of both types of figures is the same. This illusion was explained by Frith & Frith on the basis of space. They suggested that the same space appeared larger when it formed a single whole (the solid circles) than when it was divided into separate subspaces (the open circles). To test this hypothesis six new patterns were constructed. Each contained 24 circles that were divided into two groups on the basis of color. In each pattern, the circles of one color were arranged in a single linear array, and the circles of the other color in separate linear groups that were parallel to the first array. Adults and children of about eight years old were asked to estimate whether there were more circles of one color or the other using a presentation time of one second. However, the results were far from convincing. In only half of the six patterns the majority of the adults estimated that there were more circles in the single array. Results with the youngest age group seemed to be somewhat more in the predicted direction. As can be seen in Table 1 of Frith & Frith (1972).

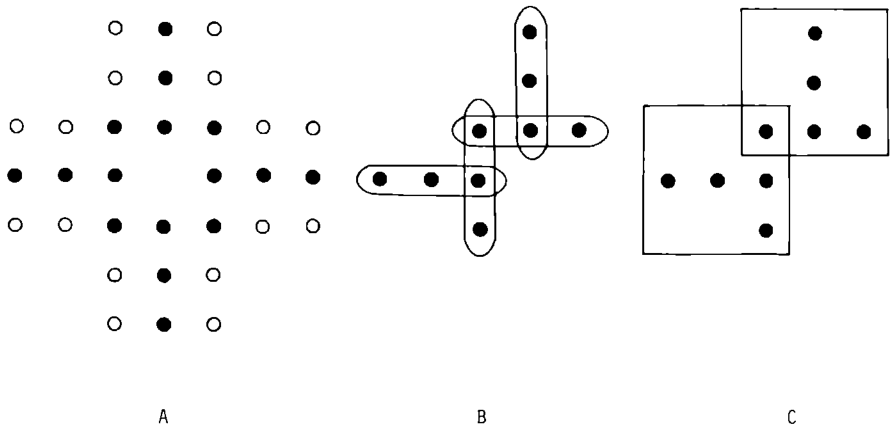


Figure 2.2. The solitaire illusion from Frith & Frith, 1972 (Figure A), a possible division of the pattern of solid figures into groups of three (Figure B) or into groups of four figures (Figure C).

The solitaire illusion may be explained in a different way than used by Frith & Frith. That is, the illusion may be the result of the manner in which groups were formed in estimating. Specifically group forming may be more problematic for the solid circles than the open ones in the pattern, because for the open figures groups are prestructured (see Figure 2.2). On the contrary, a subject must form his or her own groups within the pattern of solid figures. These figures may be subdivided into re-occurring patterns; e.g. a transformed repetition of a group of the three solid circles (Figure 2.2^{B}), or a transformation in space of a pattern of five solid circles (Figure 2.2^{C}). In the first case six solid figures are seen, and in the second five solid figures, each time against four open figures. Therefore in both cases, it appears that there are more solid than open figures.

This explanation of the solitaire illusion illustrates the possible relevancy of pattern perception in forming quantifiable groups and in subitizing. These problems will not be studied here. The present study was directed more to the question of whether subitizing is a basic process in relative quantity estimation, and if so, the age at which it begins to develop.

2.4. Research questions and hypotheses

The studies reported in the following chapters were directed at the two questions: (1) are quantifiable groups sampled in relative quantity estimation?, and (2) is the answer to the first question related to age? A prediction that may be made in regard to these questions is that if groups are sampled, then it may be expected that estimations will be affected by the ease with which quantifiable groups can be formed within a configuration.

For studying the above questions a measure is required for distinguishing configurations according to the ease by which groups can be formed. Such a measure requires some notion of the operation of subitizing in configurations consisting of numbers of objects above the subitizing range. While the operation of subitizing itself is not clearly understood, some distinctions between configurations are possible on a rational basis.

It is possible to construct configurations in which groups, consisting of a number of elements below the subitizing limit, are given. The number of elements within a group should be below the subitizing limit because each group must be quantifiable or subitizable. Although the range cannot be determined exactly, one may safely assume that a number of four or less objects can be subitized.

Various configurations can be constructed in which groups of two or four objects are randomly placed in a field. Configurations can also be constructed in which single objects are randomly placed (see Figure 2.3). Assuming that a person will try to perceive a number as large as possible, the following may be expected. Less cognitive effort is required for perceiving a number of objects in configurations of randomly placed groups of four objects than in configurations of randomly placed groups of two objects. Moreover, less effort is needed for configurations where groups of two are randomly placed than in configurations of randomly placed single objects. In the first type of configurations, groups of four are given. In the second type, to perceive a group of four, two groups of two must be identified within a context. Furthermore, the number of contexts in a completely random pattern is large, putting a greater requirement on memory than in the grouped object situation. Thus, whether or not groups are combined by addition, it may be expected that the number of elements that can be sampled within the same inspection or fixation period increases with the numerical value of the given groups within the restriction of the subitizing range.

In a relative quantity estimation task, different pairs of the three types of patterns or arrangements in Figure 2.3 may be used in various configurations. The configurations used in the present studies represented combinations of these arrangements. For example, a group 2-1 combination consisted of pairs of randomly placed objects (group 2) superimposed over single randomly placed objects (group 1). The following combinations were used: group 1-1, group 2-1, and group 4-1. Subjects were asked to estimate the most frequently occurring type of figure in various configurations composed of circles and squares.

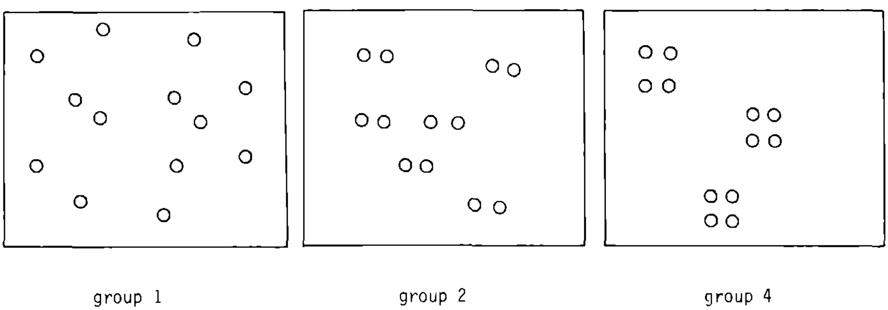


Figure 2.3. Three types of arrangement: randomly placed single (group 1), groups of two (group 2) and groups of four figures (group 4).

Using sampling by groups as the estimation strategy, one or more parts of a stimulus configuration will be inspected briefly. During an inspection a subject will try to identify subitizable groups of circles as well as of squares. It is assumed that the elements of a group are of the same type and that only a few groups can be identified. The number of elements of a group will be determined by subitizing. By performing group identification for circles as well as for squares at the same time a quantitative comparison can be made of both types of figures on each inspection. The outcomes of several inspections may be combined into an estimation by selecting the type of figure that occurred as 'more' on most of the inspections.

The number which may be obtained for circles or squares on an inspection will be affected positively by the ease with which groups of circles or squares can be identified. Generally, it should be expected that groups can be identified relatively more easily in a group 4 arrangement than in a group 2 arrangement, and in a group 2 more easily than in a group 1 arrangement. Therefore, it is expected that estimation will systematically favour the type of figure with the higher arrangement value, or group 4 above group 2, and group 2 above group 1.

Sampling by groups may be considered as a sequential process, thus requiring time. Therefore one may expect that more information will be sampled, by inspecting several times, when a configuration is presented for a longer time. Assuming that a subject is unbiased in selection of the places of sampling it may be expected that more accurate estimations can be made if more information is sampled. In order to investigate these presuppositions presentation time and proportion were used as independent variables in addition to arrangement. To investigate the accuracy of estimation independent of arrangement, different proportion values, ranging from .40 to .60, were used in the configurations. This range was used because in proportion estimation studies it has been shown that the variance of responses is largest for this range. Thus, it may be expected that even a proportion of figures of one type of .40 will be incorrectly estimated occasionally as greater than a proportion of figures of another type of .60 in a configuration. However, a proportion of .60 will be estimated more frequently higher than a proportion of .40. The relative frequency of higher estimations of proportions, ranging from .40 to .60, may be regarded as an index of accuracy.

Finally, the relationship between age, ranging from six years to adults, and estimation was explored. There is some suggestion on the basis of research discussed earlier (section 2.1) that sampling by groups begins between the ages of about five to seven. Thus the age of six was included for study here, as well as ages eight, twelve and undergraduate psychology majors. It was further predicted that as with many other performance tasks the accuracy of estimation will increase with age.

The following hypotheses were investigated:

1) *Higher estimation of the number or proportion of one type of figure in a configuration relative to another type will be obtained more frequently when this number or proportion is represented by a group 2, or a group 4 arrangement (in group 2 - group 1, or group 4 - group 1 configurations) than when it is represented by a group 1 arrangement (in group 1 - group 1 configurations). This tendency will increase from group 2 to group 4.*
2) *The frequency of higher estimations of the number or proportion of one type of figures in a configuration relative to another type will increase from .40 to .60.*
3) *The accuracy of estimation will increase with presentation time. The extent of the accuracy improvement may depend on the type of arrangements presented.*
4) *Latencies of responding will increase with longer presentation times.*
5) *The accuracy of estimation will increase with age.*

3. EXPERIMENT I

In this and the following chapters four studies will be reported which concerned sampling by groups and the development of this process. The first study, which will be reported in the present chapter, tested the hypotheses formulated at the end of the last chapter. Furthermore, it functioned as the basis for the other studies. These studies were designed to be similar in some aspects and depended on comparable principles. Methods common to all four studies will be described extensively in the method section of the first study.

3.1. Method

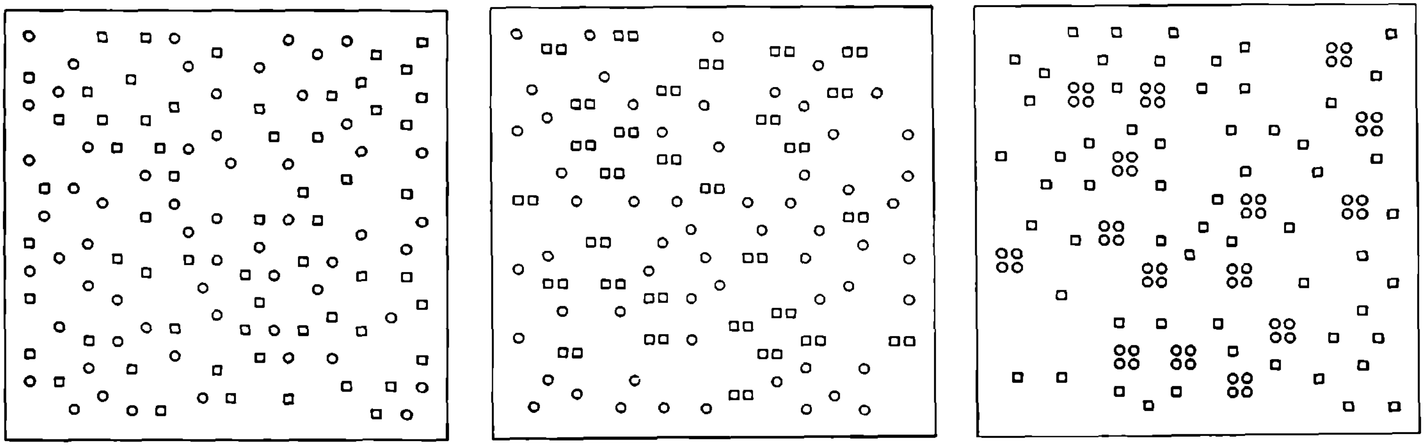
Subjects.

Twelve undergraduate psychology students at the University of Nijmegen, and 36 schoolchildren served as subjects. The group of children included 12 sixth graders with a median age of 12-3 (range 11-9 to 13-1), 12 second graders with a median age of 8-2 (range 7-5 to 8-7) from an elementary school in Nijmegen, and 12 kindergartners with a median age of 6-2 (range 5-9 to 6-3) from two nursery schools in Nijmegen. There were approximately equal numbers of boys and girls at each age level.

The undergraduates volunteered to be subjects in the experiment without payment. However, the schoolchildren and kindergartners were given comics, marbles, or feltpens afterwards, in appreciation of their participation.

Stimuli

Ninety-six stimulus configurations were constructed, each with 120 elements of two types, circles and squares. The configurations varied in proportion and arrangement of the elements. The proportion of circles or squares was .40, .50 or .60 with 48, 60 or 72 circles or squares in a stimulus configuration. At the .40



A B C

Figure 3.1. An example of three differently arranged stimulus configurations: group 1 (A), group 2 (B), and group 4 (C). The proportion of circles to squares is .50 for each configuration.

and .50 proportions there were three different types of arrangements of circles or squares (group 1, group 2, and group 4). However, at the .60 proportion only group 2 and group 4 arrangements were used. Arrangement variation was established in the following way. In some configurations, both circles and squares were placed randomly (group 1). In other configurations one of each type of figure was ordered separately and randomly placed in groups of two (group 2) or four (group 4) and the other type was arranged as in group 1, that is, randomly placed (see Figure 3.1).

For eight of the nine arrangement by proportion conditions basic patterns were constructed. By interchanging circles and squares in each basic pattern, 12 stimulus configurations were made for each condition (see Table 3.1.). Stimulus configurations for the group 1 condition were not made at a proportion of .60 to reduce the number of stimuli, and because configurations for this condition were identical to the configurations in the group 1 condition for the proportion of .40 (see Table 3.1 third row, first and fourth column).

Table 3.1. Overview of the 96 stimulus configurations. Each cell marked (+) represents half of the number of configurations in a condition, that is, six configurations (configurations were not presented for empty cells).

Proportion	ARRANGEMENTS: circles			squares		
	group 1	group 2	group 4	group 1	group 2	group 4
.40	+	+	+	+	+	+
.50	+	+	+	+	+	+
.60		+	+		+	+

The stimulus configurations were generated by a PDP-11/45 computer in the following manner: the number of elements of the two types and arrangement for them was read from a data file. The locations of the figures were drawn from a 28x28 matrix using a constraint random procedure. The 28x28 matrix proved most suitable for the generation procedure and presentation apparatus that were used (see appendix 1, flow diagram). A schematic representation of the most important constraints, those related to arrangement, is given in Figure 3.2. These arrangement constraints were necessary to get clearly separated groups of two or four figures when needed, and to prevent groups from forming when the arrangement was meant to be ungrouped. As is illustrated in Figure 3.2, in a group 1 arrangement, single elements were never located next to each other. In grouped arrangements, groups of figures were located separately from each other, and the elements of a group were placed adjacent to each other (horizontally in a group of two and square in a group of four).

```
        +                  +  +                  +  +
     -  -  -            -  -  -  -            -  -  -  -
  +  -  F  -  +      +  -  F  F  -  +      +  -  F  F  -  +
     -  -  -            -  -  -  -         +  -  F  F  -  +
        +                  +  +               -  -  -  -
                                                 +  +

        A                    B                    C
```

Figure 3.2. Arrangement constraints around an ungrouped figure (A), a group of two (B) and a group of four figures (C). The positions marked (F) denote figures of the same type (circles or squares) Given (F), the positions with (-) should be unfilled, and the positions with (+) may only be filled with figures of the other type.

There were additional constraints concerning the minimum number of elements of the same type for each quadrant of the matrix. For a proportion of .40, this minimum was 10 elements with a group 1 arrangement and 11 with an arrangement using groups of two. For the other condition the minimum number of elements was 12. These minima were determined empirically using the time in which a configuration could be generated as the criterion. The larger the minimum number, the longer the time required to find all locations of a configuration. Minimum numbers were set up to obtain regular spacing of groups over the field.

Given the constraints, locations were drawn. Depending upon the type of arrangement one location, a group of two locations, or a group of four locations were drawn. Locations for the two types of elements were drawn alternately. After the requisite number was obtained for one type, drawing of locations was continued for the remaining type. By interchanging the positions of the two types of elements, two configurations were acquired for each randomization. Six basic patterns were produced for each of the eight conditions (see Table 3.1). A file of the 96 configurations was created for computerized storage of the patterns. With this file any configuration could be generated and displayed within about 6.5 seconds.

Apparatus

Configurations were presented on video monitors controlled by computer. Once a configuration was generated by the computer, the circles and squares were displayed on two 12 inch video monitors (Philips LDA 2105) using a Scan Converter, PEP-400-R Video Graphic Storage Terminal. The circles and squares were shown as light non-solid figures against a medium gray background. The diameter of each circle was .43 cm. The area of each square was equal to that of each circle. Two monitors were used so that two subjects could participate in the experiment at the same time. Keyboards (or response boxes) were used as response apparatus. Each keyboard contained two push-buttons, one of which was placed in the midst of a square and the other in the midst of a circle. Both circle and square were solid black with a diameter of six cm. Subjects responded by pushing one of the two buttons. Responses, latencies, and omissions were recorded by computer. Latencies were measured by computer to the nearest millisecond from the time the stimulus appeared on the screen until a button was pushed. Stimuli were displayed for 1, 4, or 7 seconds. Latencies of longer than five seconds after the stimulus disappeared were treated as omissions. Omissions were displayed for each keyboard. An overview of the equipment used is given in Figure 3.3.

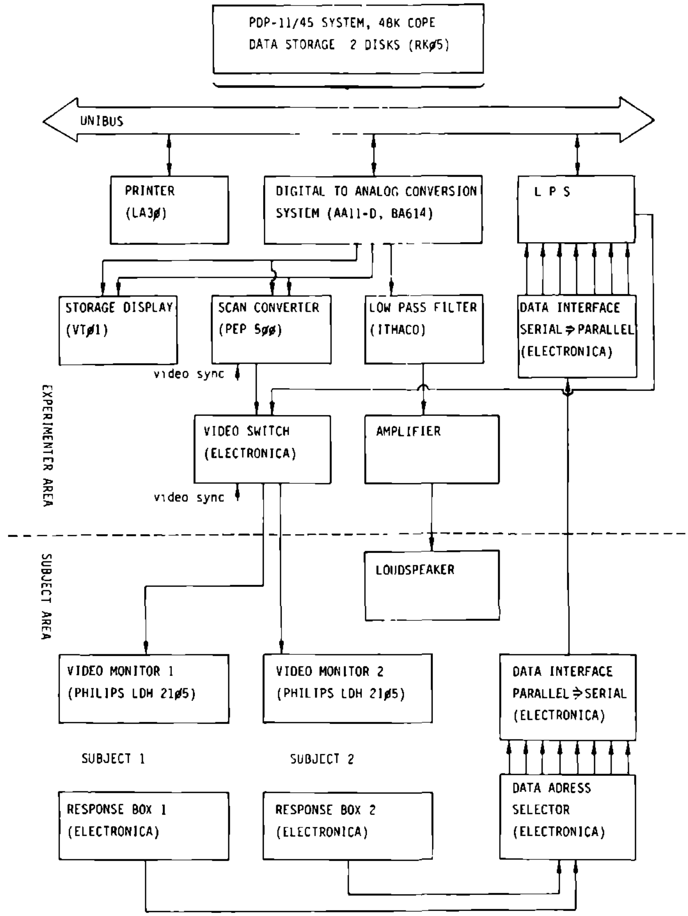


Figure 3.3. Schematic overview of apparatus used. Boxes labelled as ELECTRONICA represent specially made apparatus. The experiment was run using the single operating system RT-11F/Bc, partially with newly developed programs, partially with subroutines of the library PSYLIB (Maarse, Versteegen, & Haenen, 1976).

Procedure

Each of the 96 configurations was presented to the subjects for 1, 4 or 7 seconds. Subjects were not able to count the single elements or even the groups within the longest exposure time. To control for order effects, the 96 configurations were divided into six blocks of 16 stimulus configurations. Within every block there were two configurations for each of the eight conditions. These two stimuli contained the same basic pattern. A unique random order was prepared for presentation of each block. The presentation time for the stimuli was the same within a block but different between blocks. For all blocks inter-stimulus intervals were set at a value of seven seconds. All three exposure times were used within triads of subsequent blocks. By counterbalancing the order of the blocks and the order of presentation times over blocks, six different sequences of 288 presentations were generated.

Within each age group subjects were arbitrarily assigned to one of the six sequences. For each sequence, there were two subjects within an age group. Subjects were seen on two separate occasions with an interval of a week between sessions. The first nine blocks of a sequence were presented the first session, the remaining nine blocks the second session. Most subjects assigned to the same sequence were seen together on both occasions. On the average, the first session lasted about 50 minutes, and the second about 40 minutes.

Subjects were seated 110 cm. from a video monitor with a keyboard at a comfortable distance in front of them. The video monitor was tilted slightly backwards so that a projection from the eyes of a subject on the centre of a screen would be approximately perpendicular. It was impossible for a subject seated in position to see the other subject's monitor, or the rest of the room (Figure 3.4).

At the beginning of the first session, an instruction series and a practice series were presented, containing six configurations each. The proportions were .80 and .20 in the instruction series, and .60 and .40 in the practice series. The configurations in each of the two series represented the six combinations of the three types of arrangements and proportions. Each of the series was presented in random order. The instruction series were untimed, the practice series were timed. The instruction series was designed to clarify the directions, to familiarize subjects with the response apparatus, and to control for possible misunderstandings of the tasks and the concepts involved, especially in the youngest age group. The extreme proportion values of .80 and .20 were used because it was assumed that with them subjects could understand the instructions sufficiently to provide correct ans-

Figure 3.4. The subjects' position in experimental set up and the placing of the monitors.

wers. The criterion was a correct response for each of the six configurations in the instruction series.

During the presentation of the first configuration in the instruction series, the following was said by the experimenter: "On the television in front of you, you can see a picture with circles and squares. Look carefully and try to figure out if there are more circles or more squares. When there are more circles push the button with the large black circle close to you. When there are more squares, push the button with the large black square. Do you have any questions?" All questions were answered in terms of these directions. The terms counting and estimating were deliberately never mentioned. Once a response was given, the next configuration was displayed until the whole series had been presented.

At the two youngest age levels, the instructions were repeated with each subsequent presentation. Subjects from these two age groups were also requested to quantify immediately groups of two and of four, that is, they were asked to say how many were in the group. Besides, they were asked to point once to circles and squares on the screen. All subjects at these age levels appeared able to discriminate readily between circles and squares,and to immediately quantify groups of two and of four. These capacities were assumed at higher age levels.

All subjects responded correctly on each of the six configurations of the instruction series. Following the presentation of the instruction series, subjects were told that many of these pictures would be presented and that the pictures would be visible sometimes for a short, sometimes for a longer, and sometimes for a still longer time. They were told, furthermore, that a practice series of six pictures would be given in the course of which the first two pictures would be presented for the longest time, the second two for a shorter time and the last two for a still shorter time. They were also told that a warning signal would be given shortly before the appearance of each picture. In addition, the essence of the instructions given at the beginning of the instruction series was restated. Besides, they were asked to push a button when they knew the answer, but to respond on each presentation. They were told that responding could be done until the next configuration appeared.

The stimuli of the practice series were presented as follows: the first two for seven seconds each, the next two for four seconds each, and the last two for one second each. The inter-stimulus intervals were seven seconds for all presentations. A warning signal was given of 500 milliseconds before each configuration of the practice series and of the main series appeared on the video monitors.

After the practice series had been presented, the essential directions were again repeated. This was also done at the beginning of the second session. Subjects in the three youngest age groups were finally told that they could win a prize at the end of the second session if they did their best. Subsequently, the main series of stimuli were presented.

3.2. Results

Frequency of 'more' responses on 12 presentations was used as the estimation score for each condition. For example, the score for a proportion of .50 for group 1, was calculated by counting the total 'more' responses in six of the configurations for circles and in the six other configurations for squares. The scores for the proportion of .60 in the three group 1 presentation time conditions were derived by subtracting from the maximum score of 12, the scores for a proportion of .40 in the three comparable group 1 conditions.

The effects of the variables, proportion, arrangement and presentation time, each with three levels, were analysed across and within age groups. Across age groups, a 4(age) x 3(proportion) x 3(arrangement) x 3(presentation time), univariate analysis of variance was performed, with repeated measures on the last three factors (see Table 3.2)* Age was treated as a random factor because ages were selected rather arbitrarily with exception of age six. The repeated measurement factors were treated as fixed. Because of the differences in variance within age groups, separate repeated measures analyses of variance were performed for each age (see Tables 3.3, 3.4, 3.5, and 3.6). Since there was a dependency in the data caused by the derived scores for the .60 group 1 conditions, all effects were tested at the .01 level of statistical significance. Type I errors are kept relatively small by this low level of significance because it seemed safe in this case to increase the risk of type II errors over type I errors.

Finally, multivariate analyses of variance were used to explore age differences in more detail. A multivariate rather than a univariate approach of the repeated measures design was used because the assumptions of homogeneity of variances and covariances are not necessary with the multivariate approach. With the univariate, the structure of the covariance matrix of the repeated measures has to be of the compound symmetry pattern.

The arrangement hypothesis (hypothesis 1)

According to this hypothesis, for group 4, proportions will be estimated to be higher more frequently than for group 2 and group 1, and for group 2, more frequently than for group 1.

---

*) Kwaaitaal, T. & Roskam, E. Analysis of Variance PSYLAB VARIAN/01. Catholic University of Nijmegen, 1968. This program was used for all univariate analyses of variance.

Table 3.2.: Univariate analysis of variance for frequency of responses across age.

SOURCE	df	MS	F
Age (A)	3	546.71	6.25[a]
subjects within age	44	87.43	
Proportion (P)	2	2 477.53	14.36[a]
Arrangement (Arr)	2	225.48	1.53
Presentation time (T)	2	8.71	2.14
A x P	6	172.51	19.88[c]
A x Arr	6	147.46	5.50[b]
A x T	6	4.07	2.24
subj w x A x P	88	8.68	
subj w x A x Arr	88	26.80	
subj w x A x T	88	1.82	
P x Arr	4	33.64	2.03
P x T	4	59.29	11.69[b]
Arr x T	4	7.58	4.67
A x P x Arr	12	16.54	4.05[c]
A x P x T	12	5.07	2.47[a]
A x Arr x T	12	1.62	.99
subj w x A x P x Arr	176	4.09	
subj w x A x P x T	176	2.06	
subj w x A x Arr x T	176	1.64	
P x Arr x T	8	3.67	2.19
A x P x Arr x T	24	1.67	.86
subj w x A x P x Arr x T	352	1.95	

Note: [a] $p < .01$
[b] $p < .001$
[c] $p < .001$

Table 3.3.: Univariate analysis of variance for frequency of responses for adults.

SOURCE	df	MS	F
within subjects	11	20.05	
Proportion (P)	2	1651.75	267.67[3]
Arrangement (Arr)	2	67.29	6.65[1]
Presentation time (T)	2	17.58	15.08[2]
subj x P	22	6.17	
subj x Arr	22	10.11	
subj x T	22	1.16	
P x Arr	4	41.31	10.74[3]
P x T	4	35.53	15.43[3]
Arr x T	4	5.23	2.95
subj x P x Arr	44	3.85	
subj x P x T	44	2.30	
subj x Arr x T	44	1.77	
P x Arr x T	8	1.75	.87
subj x P x Arr x T	88	2.00	

[1] $p < .01$
[2] $p < .001$
[3] $p < .0001$

Table 3.4.: Univariate analysis of variance for frequency of responses for 12 year old children.

SOURCE	df	MS	F
within subjects	11	35.52	
Proportion (P)	2	891.73	77.62[3]
Arrangement (Arr)	2	462.38	36.40[3]
Presentation time (T)	2	1.43	
subj x P	22	11.49	
subj x Arr	22	12.70	
subj x T	22	1.93	
P x Arr	4	32.94	6.22[2]
P x T	4	28.70	15.47[3]
Arr x T	4	5.30	3.33
subj x P x Arr	44	5.30	
subj x P x T	44	1.86	
subj x Arr x T	44	1.59	
P x Arr x T	8	3.07	1.69
subj x P x Arr x T	88	1.82	

[2] $p < .001$
[3] $p < .0001$

Table 3.5.. Univariate analysis of variance for frequency of responses for eight year old children.

SOURCE	df	MS	F
within subjects	11	182.20	
Proportion (P)	2	320.46	30.61[3]
Arrangement (Arr)	2	73 24	1.50
Presentation time (T)	2	.20	.13
subj x P	22	10.47	
subj x Arr	22	48.72	
subj x T	22	1.54	
P x Arr	4	5.40	1.33
P x T	4	5.61	3.28
Arr x T	4	1.38	.88
subj x P x Arr	44	4.07	
subj x P x T	44	1.71	
subj x Arr x T	44	1.57	
P x Arr x T	8	1.85	.89
subj x P x Arr x T	88	2.08	

[3] $p < .0001$

Table 3.6.: Univariate analysis of variance for frequency of responses for six year old children.

SOURCE	df	MS	F
within subjects	11	111.92	
Proportion (P)	2	131.07	19.89[3]
Arrangement (Arr)	2	64.90	1.82
Presentation time (T)	2	1.64	.62
subj x P	22	6.59	
subj x Arr	22	35.69	
subj x T	22	2.64	
P x Arr	4	3.63	1.15
P x T	4	4.69	1.98
Arr x T	4	.59	.35
subj x P x Arr	44	3.15	
subj x P x T	44	2.36	
subj x Arr x T	44	1.63	
P x Arr x T	8	2.02	1.06
subj x P x Arr x T	88	1.90	

[3] $p < .0001$

The analysis of variance across age groups yielded a non-significant main effect for arrangement, $F(2,6)=1.53$, n.s. (see Table 3.2). However, the interaction between age and arrangement was significant, $F(6,88)=5.50$, $p < .001$, indicating that for at least some ages arrangement significantly affected estimation. Analyses of variance within age groups revealed a significant arrangement effect in adults, $F(2,22)=6.65$, $p < .01$, and in children of 12 years old, $F(2,22)=36,40$, $p < .0001$, but, unexpectedly, not in children of eight and six years old (see Tables 3.3, 3.4, 3.5, and 3.6). Figure 3.5 shows that subjects in the two oldest age groups, as expected, estimated group 2 and group 4 proportions more frequently higher than group 1 proportions.

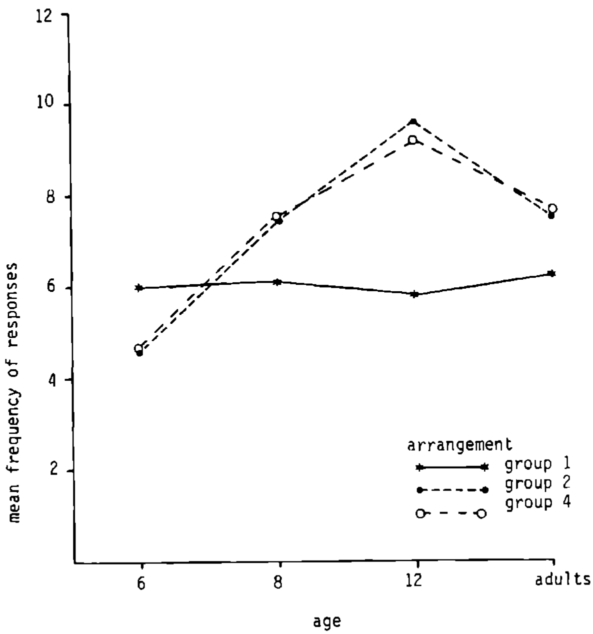


Figure 3.5.. Interaction between age and arrangement.

The analysis across age groups revealed a significant interaction between age, arrangement and proportion, $F(12,176)=4.05$, $p < .0001$. The arrangement effect in adults and children of age 12 appeared to be dependent on proportion (for adults, $F(4,44)=10.74$, $p < .0001$, and for 12 year olds $F(4,44)=6.22$, $p < .001$, see Figure 3.6).

Newman-Keuls analyses at the .01 level (Winer, 1971, p. 442-443) revealed that group 4 and group 2 proportions did not significantly differ from each other, except at the .50 proportion value in adults and at the .40 value in 12 year olds. Only in adults, at the propoition value of .50, the predicted difference between group 4 and group 2 was found. At that value the frequency of higher estimations was lower for group 2 than for group 4. However, the reverse effect was found at the value of .40 in children of age 12. So, the arrangement hypothesis was partially supported in adults and children of 12 years old. However, in eight and six year old children this hypothesis was not confirmed.

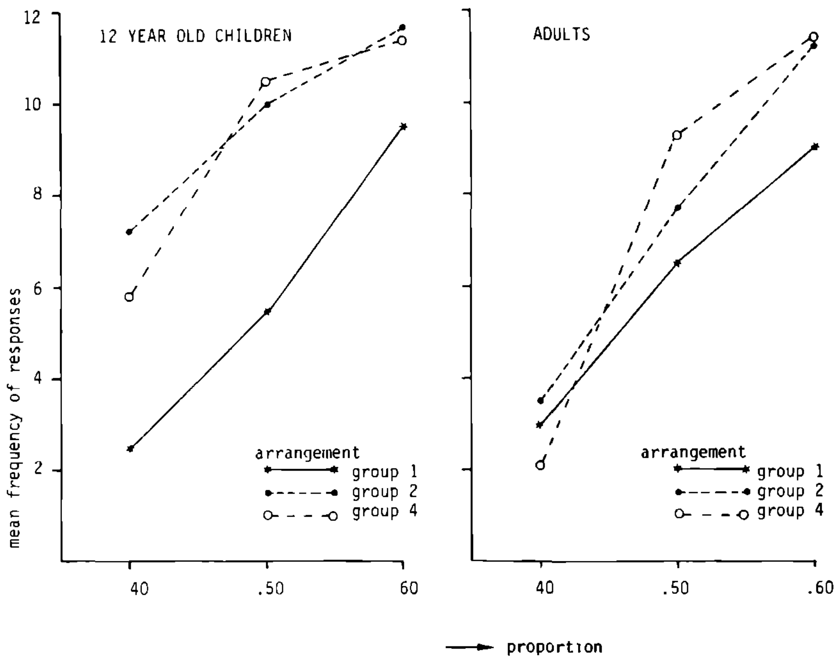


Figure 3.6. Interaction between arrangement and proportion within two age groups.

The analysis of variance within age groups revealed nearly significant interactions between arrangement and presentation time, in adults, $F(4,44)=2.95$, $p < .03$, and in children of 12 years old, $F(4,44)=3.33$, $p < .02$, indicating that the arrangement effect tended to decrease with presentation time, especially in adults (see Figure 3.7).

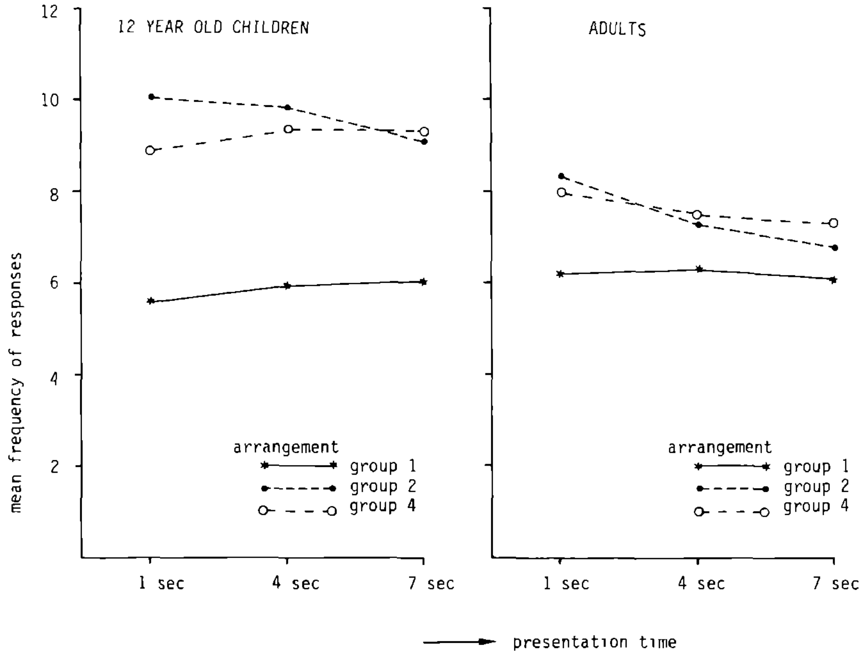


Figure 3.7. Interaction between arrangement and presentation time within two age groups.

The accuracy and presentation time hypotheses (hypotheses 2, 3, 4, and 5).

According to these hypotheses, higher proportions will be estimated as higher more frequently than lower values (hypothesis 2). Furthermore, accuracy will increase with presentation time (hypothesis 3) and with age (hypothesis 5). Latencies of responding will be longer for longer presentation times (hypothesis 4).

The analysis across age groups yielded significant effects of proportion, $F(2,6)=14.36$, $p < .01$, proportion x age, $F(6,88)=19.88$, $p < .0001$ and proportion x age x presentation time, $F(12,176)=2.47$, $p < .01$. The analyses within age groups showed significant proportion effects at all ages suggesting that higher proportion values were estimated more frequently as higher than lower values (in adults, $F(2,22)=267.67$, $p < .0001$; at the age of 12, $F(2,22)=77.62$, $p < .0001$; at the age of eight, $F(2,22)=30.61$, $p < .0001$; at the age of six, $F(2,22)=19.89$, $p < .0001$. So, hypothesis 2 was confirmed at all ages. Moreover, accuracy increased with age supporting hypothesis 5 (see Figure 3.8). Besides, for

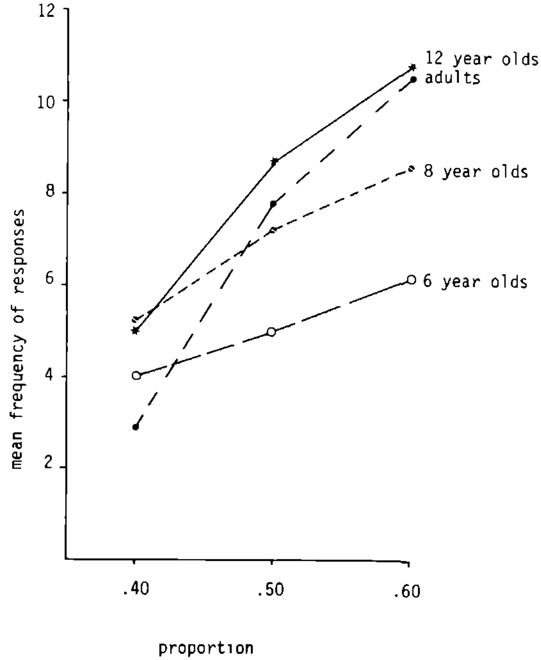


Figure 3.8. Interaction between age and proportion.

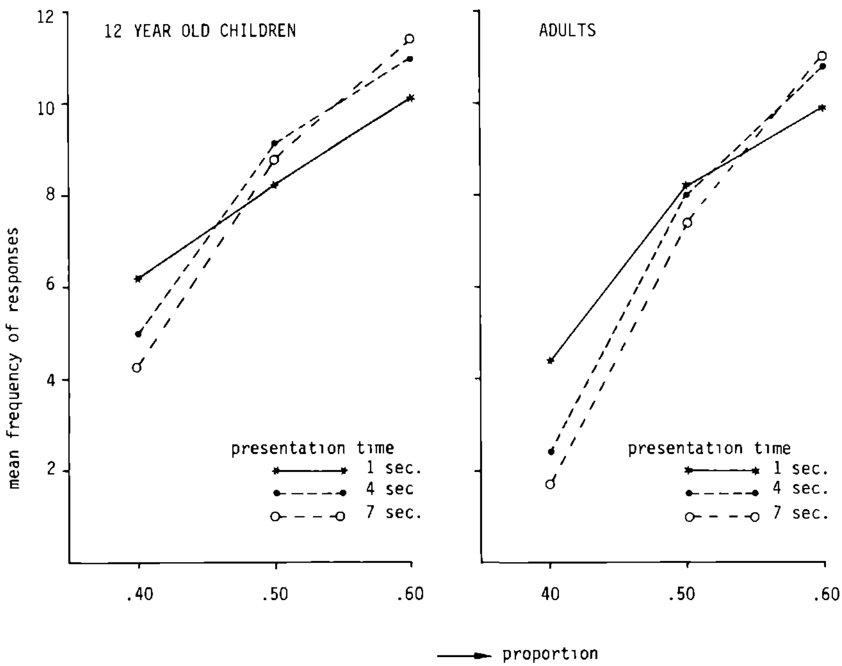


Figure 3.9. Interaction between proportion and presentation time within two age groups.

adults and 12 year olds, accuracy increased with presentation time independent of arrangement (in adults, $F(4,44)=15.43$, $p < .0001$, and in 12 year olds, $F(4,44)=15.47$, $p < .0001$, see Figure 3.9). So hypothesis 3 was confirmed for the two oldest age groups, but not for the two youngest.

Within each age group, the effect of presentation time on latency averaged across proportions, arrangements, and the 12 presentations within each condition, was analyzed. Latencies in the .60 condition were omitted, because of the missing scores in some of these conditions. Analyses of variance showed a significant presentation time effect within each age group indicating higher latencies with longer presentation times (in adults, $F(2,22)=173.96$, $p < .0001$; in 12 year olds, $F(2,22)=50.91$, $p < .001$; in eight year olds, $F(2,22)=65.27$, $p < .0001$; and in six year olds, $F(2,222)=43.75$, $p < .0002$). The mean latencies for each presentation time within each age group are represented in Figure 3.10. The large differences in latency between age groups, especially at the presentation time of 7 seconds were noteworthy. Subjects at the age of six or eight years used less than half the time adults took at this presentation time.

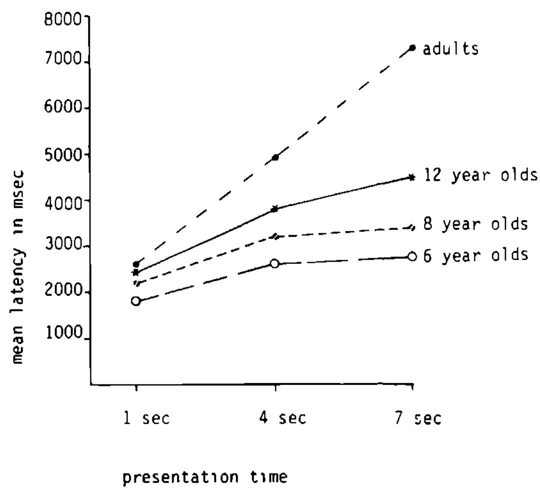


Figure 3.10. Increase in mean latency with presentation time within four different age groups.

Exploration of age differences.

In the analyses described thus far, all three factors, arrangement, proportion, and presentation time differentially affected subjects' estimations, depending on their age. Further, the mean latency of responding increased with age, especially at the seven second presentation time. It is possible that the differences in estimation between age groups resulted from a latency difference; that is, younger subjects may have explored the presented stimulus configurations in the same way as older subjects, but less extensively. The absence of a significant proportion x presentation time interaction in the two youngest age groups, and the increase of accuracy with age might be explained in this way. On the other hand, the absence of an arrangement effect in the two youngest age groups is not easy to reconcile with this point of view because in the two oldest age groups the effect of arrangement seemed to decrease with presentation time. Despite these divergent results, it is important to examine whether age differences in estimation might be explained in terms of latency of responding.

To test this hypothesis a posteriori, multivariate analyses of variance were performed on estimations and latencies separately, and in combination with each other, at the seven second presentation time. Because of the missing scores in one of the .60 proportion conditions, and the complexity of the analyses the proportion factor was reduced to two levels .40 and .50, and the arrangement factor to group 1 and group 2. The following orthogonal comparisons or contrasts were made on the repeated measures: the multivariate grand mean across all contrasts was compared to zero; .40 was contrasted to .50; group 1 to group 2; and the proportion x arrangement interaction was also tested.

First, general level of latency was tested as a covariate in age comparisons on estimation. That is, it was examined whether arrangement and proportion differentially affected latency according to age, and otherwise, if these factors affected latency pooled across age.

The multivariate comparison between all age groups on the repeated measures contrasts revealed significant differences in latency, F multivariate $(12,108)=5.95$, $p < .0001$. However, significant age differences were not found on the arrangement, proportion and arrangement x proportion contrasts, but only on the general level contrast, F univariate $(3,44)=30.30$, $p < .0001$. A multivariate analysis across age yielded significant repeated measures contrasts, F multivariate $(3,42)=9.84$, $p < .0001$. The only factor that affected latency appeared to be arrangement, F univariate $(1,44)=21.64$, $p < .0001$. However, this effect disappeared when both proportion and general level were included, the F stepdown $(1,44)=.32$. However, the effect remained significant on proportion alone, F stepdown $(1,44)=22.98$, $p < .0001$. Therefore, the arrangement effect may be attributed to differences between subjects in general level of latency, suggesting that general level of latency may be useful as covariable.

Furthermore, a multivariate comparison between age groups was performed on estimation, and yielded significant age differences F multivariate $(12,108)=4.30$, $p < .0001$. Arrangement, F univariate $(3,44)=4.34$, $p < .01$, and proportion, F univariate $(3,44)=9.83$, $p < .0001$, affected estimation differentially according to age. In this comparison, when general level of estimation was substituted by general level of latency, and the contrasts of proportion and arrangement were tested conditional on latency, the differential effects remained significant: for proportion, F stepdown $(3,44)=4.04$, $p < .01$, for arrangement, F stepdown $(3,44)=5.80$, $p < .0025$. Therefore, age differences in estimation can not be attributed to differences in latency.

The univariate analyses showed the largest discrepancies between the two oldest age groups on the one hand and the two youngest age groups on the other. Therefore, in the following multivariate analyses 12 year old children were compared with adults, and eight year olds with six year olds, and finally 12 year olds and adults on the one hand with six and eight year olds on the other. Differential effects according to age were tested controlling for general level of latency with the frequency of 'more' responses as the dependent variable.

The multivariate comparison between adults and 12 year olds was significant, F multivariate $(4,41)=7.39$, $p < .001$. Arrangement differentially affected estimation in the univariate tests, F univariate $(1,44)=6.90$, $p < .01$. In the stepdown testing this differential effect disappeared, F stepdown $(1,44)=1.67$, n.s.

Table 3.7.: Multivariate comparison of adults and 12 year old children (F multivar. $(4,41)=7.39$, $p < .001$).

CONTRASTS	df	F UNIVARIATE	F STEPDOWN
general level of latency	1,44	26.39[3]	26.39[3]
proportion	1,44	.03	1.19
arrangement	1,44	6.90[1]	1.67
proportion x arrangement	1,44	.84	.26

[1] $p < .01$
[2] $p < .001$
[3] $p < .0001$

Table 3.8.: Multivariate comparison of two age groups: eight and six year old children (F multivar. $(4,41)=2.52$, n.s.).

CONTRASTS	df	F UNIVARIATE	F STEPDOWN
general level of latency	1,44	2.29	2.29
proportion	1,44	4.01[1]	2.69
arrangement	1,44	2.36	4.72[1]
proportion x arrangement	1,44	.13	.09

[1] $p < .05$

Tabel 3.9.: Multivariate comparison of two pooled age groups: adults + 12 year old children, and six + eight year old children (F multivar. $(4,41)=22.49$, $p < .0001$).

CONTRASTS	df	F UNIVARIATE	F STEPDOWN
general level of latency	1,44	62.23[3]	62.23[3]
proportion	1,44	25.45[3]	4.45[1]
arrangement	1,44	3.77	8.17[2]
proportion x arrangement	1,44	.68	

[1] $p < .05$
[2] $p < .01$
[3] $p < .0001$

(see Table 3.7). Further, general level of latency differed for adult and 12 year old subjects, F univariate =26.39, $p < .0001$, indicating a longer mean latency of responding for the seven second presentation time in adults than in 12 year old children. The multivariate comparison between six and eight year old subjects yielded no significant effects, F multivariate $(4,41)=2.52$, n.s. (see Table 3.8). However, the comparison between the two older age groups and the two younger age groups revealed significant effects, F multivariate $(4,41)=22.39$, $p < .0001$. The effects of proportion and arrangement were significant, respectively, F stepdown $(1,44)=4.45$, $p < .05$, and F stepdown $(1,44)=8.17$, $p < .01$ (see Table 3.9).

3.3. Discussion and conclusions

The findings of the present study, particularly in adults and in 12 year old children, offer support for the hypothesis that estimating is done by sampling by groups. The predicted arrangement effect was found for group 2 and group 4 arrangements. Furthermore, in these age groups the predicted increase of accuracy with presentation time was found. The absence of a larger effect for group 4 compared with group 2 arrangements, except for a proportion of .50 in adults, seems to be rather puzzling at first sight. The absence of similar differences in the proportion of .60 may be explained easily in terms of ceiling effects. However, another explanation is needed for the lower proportion values. In fact at a proportion of .40 in adults, neither a difference between group 4 and group 2, nor an arrangement effect was found.

A combination of factors that may have been partially caused by the type of stimulus material and task used, may be responsible for these unexpected findings. First, in group 2 configurations, all groups were placed horizontally. In that situation, only slightly more cognitive effort may be required to perceive a group of 4, or to perceive and add two groups of 2, than to perceive a square formed by a group of 4 in the group 4 arrangements. Second, in the sampling by groups hypothesis the acquisition of some numerical value for both types of figures was assumed each time a part of a stimulus configuration is fixated. However, with respect to the task used some modification of this assumption seems needed in retrospect. Suppose that four circles forming a square group are perceived adjacent to a lot of randomly placed squares. For perceiving more squares it may be sufficient to know only that there are four circles. Some results of Atkinson's et al (1976[a]) study suggest that for a comparison with a number of four a majority of six is enough. Especially for a configuration with a proportion of .40 for the grouped category, one may expect that in some parts of it, for example, four circles forming a square are located adjacent to six or more ungrouped squares. In that case it may not be necessary to acquire a numerical value for the squares to see that there are more then four of them. This situation will be more likely for configurations with a proportion of .40 than with a proportion of .50 for the grouped category. Therefore, a proportion of .40 probably is the lower limit of measuring arrangement effects using the type of task of the present study. Third, placing figures in separate groups of four leads to relatively more clustering of the other group 1 figures than placing them in separate groups of two, especially at a proportion of .40 for the grouped figures. A higher clustering may create situations where adjacent to grouped figures, a larger number of ungrouped figures is located, sufficient to perceive its majority without knowing the real number of them. This perhaps explains the relatively lower score, in some cases, for group 4, compared with group 2 arrangements at a proportion of .40.

One might interpret the arrangement effect as estimating on the basis of density because the density of figures placed in groups of two or four was higher than that of the single placed ones. If such were the case, an arrangement effect and also a larger effect for group 4 than for group 2 situations should have been found because the density was highest in the group 4 configurations. Furthermore, the effects of arrangement at different ages were clearly at odds with what might be expected from a density hypothesis. It is hardly conceivable that from the age of about 12 density is used as a cue, that is, it has to be explained why this cue was not used by the two youngest age groups, and why the use of it decreased from the age of 12. Perhaps, in the youngest age group estimations were based on visual area rather than density. Furthermore, from the age of six, perhaps, there was a developmental trend indicating a substitution of visual area by density. Such a trend would probably agree with the age x accuracy interaction that was found but not with the increase of accuracy with age. The latter trend may be hardly expected from substitution of misleading visual cues by another as invalid one like density. Furthermore, the reports of subjects in both oldest age groups were also at odds with the foregoing hypothesis. According to these reports, an estimation was based on a few number comparisons, fixating different parts of a configuration, and in group 2 and group 4 situations identifying given groups in combination with adjacent group 1 figures.

Differences between adult and 12 year old subjects could be attributed to a difference in latency of responding, a result supporting the hypothesis that the same type of process was used by subjects in both age groups. However, time as such does not explain why in adults the arrangement effect was lower and accuracy higher. One possibility is that adults screened more parts of a configuration at the seven second presentation time than 12 year old children. A more plausible alternative seems to be that inspected parts were screened more carefully. For, screening of just some more parts with presentation time would have resulted in an increase in the effect of arrangement instead of a decrease. In that case a longer presentation time would result in more replications with possibly the one second presentation having the lowest arrangement effect. That is, if just more biased samples were drawn the biased orientation would be strengthened. The results of the conditional testing procedure further suggest an absence of age differences at the one second presentation time. However, also at this presentation time the effects of arrangement and proportion appeared to differ for both age groups, F multivariate $(6,28)=3.06$, $p < .02$; significant differences were found for proportion (.40 vs. .50), F univariate $(1,33)=12.48$, $p < .01$, and for arrangement (1 vs. 2), F univariate $(1,33)=5.77$, p. < 02, indicating a lower arrangement effect, and a higher accuracy in adults. From this, one might conclude that at the shortest presentation time, adults screened more carefully than 12 year old children. This conclusion, however, seems to be challengeable. A complication is that each subject perceived the stimulus configurations for all three presentation times. It is possible that subjects profited in the shortest exposure time from the longest, and that adults profited more than 12 year old subjects. For more definite conclusions about the two oldest age groups a replication of the present study with only the one second presentation time condition is needed.

The findings within both youngest age groups did not support the arrangement hypothesis. Furthermore, unexpectedly, accuracy did not increase with presentation time. Therefore, the hypothesized way of estimating, found in both oldest age groups, may not be attributed to the eight and six year old subjects. Further support for this conclusion came from the results of the multivariate comparison between both youngest age groups on the one hand and both oldest on the other hand. Differences in the effects of arrangement and proportion could not be attributed to a difference in latency of responding as appeared to be the case with 12 year olds compared with adult subjects.

It may be concluded that the present study suggest that there may be a change with age in the way of estimating, possibly between the age of eight and 12 years. Two important questions are: (1) what type of process was used by six and eight year old children, and (2) why did they not estimate in the same way as the older subjects. The present study was not designed to answer both questions. However, some indications with respect to the first question came from the reports of the children, especially those in the youngest age group. They argued that generally there were more group 1 than grouped figures, because the first were seen everywhere in a configuration. On the other hand, grouped figures were close together, and therefore less in number. This type of arguments suggests a way of estimating that is based on visual area. However, following such a strategy, a lower score for group 4 and 2, compared with group 1 arrangements might be expected. Such a difference was not found at the age of eight and not significantly at the age of six. Possibly a visual area strategy was conflicting with attempts to estimate on the basis of perceived numbers. A comparable hypothesis was given by Flavell (1977) for the fact that a memory strategy could be taught, but was not used spontaneously at some ages. This suggestion may also explain why these children did not estimate in the same way as older subjects. It may be assumed that their number knowledge was sufficient to do so. A number conservation task (Rothenberg, 1969) given to eight of the 12 subjects of the youngest age group revealed number conservation for all tested children. Further study is needed, of course, for finding the real answers to the stated questions.

4. FURTHER EXPERIMENTS

In this chapter three follow-up experiments will be reported. Experiment II examined whether the arrangement effect could be induced by training eight year old subjects to perform the hypothesized method of estimating used by older subjects. As will be shown, this study and the fourth one are especially relevant to the conclusions drawn in the first study about age differences, and about the estimation process of children who are at least 12 years old. Experiment III was based on results of the first two experiments, and investigated differences between subjects, ranging in age from eight years to adulthood, using the one second presentation time of the first experiment. Experiment IV was an extension of the second experiment. It investigated whether the arrangement effect could be induced by training in six year old children.

4.1. Experiment II: a training study

4.1.1. Research questions and hypotheses

In the previous study it was suggested that the process of estimating used by six and eight year olds might be different from that of 12 year olds and adults. That is, sampling by groups probably was not used as method of estimation by subjects belonging to the two younger age groups. This assertion, however, rested on the assumption that the basis for the arrangement effect is sampling by groups. The second study was designed to investigate the validity of this assumption. The absence of the predicted arrangement effect in children of about eight years or younger, offers a possibility for examination of the validity of the assumption that this effect would be implied by using sampling by groups. Its validity would be strongly supported if the arrangement effect could be induced by teaching eight or six year olds this method of estimation. Furthermore, it is important to show that teaching older subjects the same method does not affect their estimating, because this taught method should by synonymous with what they already do.

In this study, a short estimation training session was administered to eight and 12 year old subjects. After that, the stimulus material used in the first experiment was presented at a constant exposure time of seven seconds per stimulus configuration. Eight year old subjects rather than younger ones were used, because these children were thought to be more easily trainable. The training procedure consisted of iteratively performing numerical comparisons with small numbers of circles and squares for different parts of a stimulus configuration, and of determining the type of figure with the higher number on most comparisons made. Although alternative specifications of sampling by groups may be possible in principle, it was assumed that this procedure most closely represented what 12 year old and older subjects were doing in the previous experiment.

Training effects were investigated on the basis of estimations and latencies. From earlier results it was expected that training would increase latencies of eight year olds, because estimating according to the trained method seems to be more time consuming than according to the strategy common for that age. This effect, however, could hardly be expected in 12 year olds unless unintentionally, by training they will estimate more carefully. That is, by training, 12 year old subjects may begin to estimate more like adults showing an increase in accuracy and a decrement of the arrangement effect.

Latencies may also be affected by arrangement when the trained method of estimation is applied. That is, group identification may be done more easily in group 4 than in group 2, and in group 2 more easily than in group 1 situations. However, this effect may be confused by other factors, for example, the number of comparisons made, care in judging, and the specific part of a configuration that is inspected. Despite the recognition of possible confounding by other factors, arrangement effects on latency were explored in this study. These effects may be indicative of whether or not the procedure that they were trained to use would be applied by subjects during estimation.

The following three hypotheses were tested:

1) *At the age of 12, group 2 and group 4 proportions will be estimated more frequently higher than comparable group 1 proportions by 12 year old subjects. This arrangement effect will be independent of training and may be dependent on the specific proportion value.*
2) *At the age of eight, group 2 as well as group 4 proportions will be estimated higher more frequently than comparable group 1 proportions only by trained eight year old subjects. This arrangement effect may be dependent on the specific proportion value.*
3) *By training eight year old subjects, their latency in responding will increase.*

4.1.2. Method

Subjects

The subjects were 24 second-, and 24 sixth grade children from an elementary school in Nijmegen. There were an equal number of boys and girls at each grade level. Within every grade childrer were assigned in equal numbers to the control or the training group. Assignment was random with the constraint that the groups be balanced for sex. The median age of each group of subjects at each grade level was as follows:

	Training group		Control group	
	Range	Median	Range	Median
age 8:	7-3 to 7-11	7-6	7-1 to 8-1	7-6
age 12:	10-9 to 12- 8	11-7	11-6 to 13-3	12

All children were given comicbooks in appreciation for their participation.

Procedures

The 96 stimulus configurations of the first experiment, and 12 new ones, representing the group 1 .60 proportion condition (six with a .60 proportion of circles, and six with a .60 proportion of squares), were presented for seven seconds each. The whole series of 108 stimulus configurations was divided into six blocks of 18 stimuli, and presented in one session, counterbalancing for the order of blocks over subjects. Blocks were formed using the six blocks of 16 stimuli of the first experiment, and adding two of the new stimulus configurations to each block. Apparatus, instruction, and procedure were the same as in the first experiment except for the estimation training.

The estimation training was given using a training series containing 12 stimulus configurations that had been constructed by computer. Each was printed on a separate piece of paper. The first 10 stimulus configurations of the training series consisted of 18 or 28 figures (circles and squares), and the last two of 120 figures, the number used in the main series. The number of figures in the first 10 stimuli was kept relatively small to make training easier. That is, application of the procedures to configurations might be facilitated by showing a child that the same answers should be given whether they were from estimation or counting. Therefore, only an unequal proportion of circles or squares, .40 or .60, was shown in the training series in different stimulus configurations. Further, for both proportions circles as well as squares were arranged in group 1, group 2 and group 4 configurations.

Training was given in applying the following principles:

1) *Identification of a few small parts in a stimulus configuration, each part consisting of a number of circles and squares small enough to be apprehended immediately.*
2) *For each part, determine the type of figure that occurs more frequently and call that figure 'the winner'.*
3) *Determine the figure which is the winner in most identified parts, and that figure is called 'the final winner'.*

On presentation of the first training series stimulus configuration, the subject was asked if there were more circles or squares.*) After an answer was given, he or she was asked how one could know which one was more. The importance of counting was stressed, but the impossibility of it in some situations mentioned.

---

*) The specific wording for the instructions in Dutch may be obtained by writing to the author at the Dept. of Developm. Psychol., Psychol. Lab., Catholic University of Nijmegen.

Estimation was introduced to the subjects as an alternative for counting in these situations. Estimation was explained by showing how the three estimation principles worked with the second stimulus configuration. Not only the operation of the principles was shown, but also the adequacy of it by demonstrating that counting should have resulted in the same answer. Estimation was then practiced using the rest of the stimulus configurations of the training series. That is, identification of parts was done first in a very concrete way. The subject encircled with red ink some parts in stimulus configurations 3-6, and 11. However, in stimulus configurations 7-10, and 12 identification of parts was first done mentally within seven seconds. After an estimation had been given the procedure was repeated using the concrete way, that is, identification of parts was done by encircling. Repetition was done in this form to control for correct application of the first estimation principle. As soon as about four parts were identified a numerical comparison was given by a subject for each part, followed by an estimation for the whole stimulus configuration. Comparisons were recorded by the experimenter in view of the subject. Given these comparisons the experimenter ensured correct application by a subject of estimation principle 3. That is, another answer was asked from the subject if his or her 'final winner' appeared to be not the figure that most frequently 'won' in his or her comparisons. The estimation training lasted about 20 minutes and was given to a subject individually. All subjects who received training were able to perform the estimation procedure correctly and fluently afterwards.

Before presentation of the main series, the same instructions as in the first experiment were given to all subjects using the training series as instruction series. Furthermore, for trained subjects the three estimation principles were repeated, and the instruction was given to apply these learned principles to the video situation. The training series was followed by the practice series of the first experiment and finally by the main series. For each training group subject the whole experiment required one session of about 60 minutes, and for each control group subject one of about 40 minutes. However, subjects of the latter group were given easy puzzles to solve for about 20 minutes just before the experiment started to balance the time element. All sessions were run in the morning.

4.1.3. Results

Scores were calculated in the same way as was done in the earlier experiment, except for the group 1 proportion of .60. Independent scores could be calculated now also for this condition. There were two types of scores for each condition: frequency scores, and mean latencies averaged across the 12 presentations within each condition. Multivariate analyses of variance were performed on each of both types of scores separately. In these analyses the following nine orthogonal contrasts formed the repeated measures: (1) the multivariate grand mean across all contrasts was compared to zero; (2) .40 was contrasted to .50 and (3) .50 to .60; (4) group 1 was contrasted to group 2, and (5) to group 4; furthermore the interactions were tested, (6) between the contrasts 2 and 4, (7) between the contrasts 2 and 5, (8) between the contrasts 3 and 4, and (9) between the contrasts 3 and 5. All effects were tested at the .05 level of significance.* A summary of the analyses for frequency scores is given in tables 4.1 and 4.2, and for latencies in tables 4.3 and 4.4.

Table 4.1. Summary of five separate multivariate analyses of variance of frequency scores using contrasts as the dependent variables.

Analysis	F MULTIVAR.	df	p
Trained vs untrained 12 year olds	.76	9,14	n.s.
All 12 year olds	1294.97	9,14	.0001
All 12 year olds vs trained 8 year olds	2.99	9,36	.01
Trained vs untrained 8 year olds	.36	9,14	n.s.
All 8 year olds	155.15	9,14	.0001

*) In this and following studies testing was performed less conservatively than in the first one because contrary to that study, the dependency in the data did not exist any longer.

Table 4.2.: Univariate orthogonal contrasts among means for frequency scores

Contrast	Comparisons								
	all eight year olds[a]			all 12 year olds[a]			trained eight year olds vs all 12 year olds[b]		
	MS	F	p	MS	F	p	MS	F	p
Grand mean	9814.50	302.15	.0001	12467.02	614.09	.0001	40.00	1.52	n.s.
Proportion (.40 vs .50)	655.15	93.80	.0001	1989.18	247.35	.0001	139.37	18.55	.0001
Proportion (.50 vs .60)	196.00	59.36	.0001	258.67	63.75	.0001	2.84	.77	n.s.
Arrangement (group 1 vs group 2)	36.17	8.87	.007	30.08	8.66	.008	.00	.00	n.s.
Arrangement (group 1 vs group 4)	60.06	3.98	n.s.	144.00	14.67	.001	.93	.07	n.s.
Prop (.40 vs .50) x Arr (1 vs 2)	1.11	.41	n.s.	.09	.03	n.s.	.93	.32	n.s.
Prop (.40 vs .50) x Arr (1 vs 4)	22.78	7.40	.01	2.53	.50	n.s.	7.97	1.95	n.s.
Prop (.50 vs .60) x Arr (1 vs 2)	.03	.02	n.s.	.00	.00	n.s.	.14	.09	n.s.
Prop (.50 vs .60) x Arr (1 vs 4)	6.57	1.80	n.s.	23.01	13.54	.001	4.75	1.79	n.s.

[a] error degrees of freedom were 22

[b] error degrees of freedom were 44

The effects of arrangement and training with 12 and eight year olds for estimation

According to hypothesis 1, group 2 and group 4 proportions would be estimated to be higher more frequently than group 1 proportions, independent of training for 12 year olds. To test this hypothesis a multivariate comparison was performed between trained and untrained 12 year olds on the repeated measures contrasts for frequency scores. Both groups of subjects did not differ significantly from each other, F multivariate $(9,14)=.76$, n.s. The univariate tests indicated that neither arrangement nor any of the repeated measures contrasts was differentially affected by training.

The multivariate analysis on the repeated measures contrasts across both groups of subjects produced significant results, F multivariate $(9,14)=1294.97$, $p < .0001$. The contrast between group 2 and group 1 was significant, $F(1,22)=8.66$, $p < .01$, as well as that between group 4 and group 1, $F(1,22)=14.67$, $p < .001$, indicating higher mean scores for group 4 and group 2 than for group 1 conditions (see Figure 4.1). Furthermore, the contrast between the .40 and the .50 proportion appeared to be significant, $F(1,22)=247.35$, $p < .001$, as well as the one between the .50 and the .60 proportion, $F(1,22)=63.75$, $p < .0001$, together indicating the higher the mean score the higher the proportion or $(.40 < .50 < .60)$. Finally, only the interaction between the .50 and .60 proportion, and group 1 and group 4 appeared to be significant, $F(1,22)=13.54$, $p < .001$, possibly suggesting a ceiling effect (see Figure 4.1). Thus, hypothesis 1 was supported by the results.

For testing the effect of training at the age of eight (hypothesis 2), trained eight year old subjects were compared first with all 12 year olds and then with untrained eight year old subjects. Finally, the repeated measures contrasts were tested for all eight year olds.

A multivariate analysis revealed that trained eight year olds differed significantly from 12 year olds, F multivariate $(9,36)=2.99$, $p < .01$ on the repeated measures. Univariate testing showed that both groups differed significantly only for the .40 proportion compared with .50, $F(1,44)=18.55$, $p < .0001$. Thus, trained

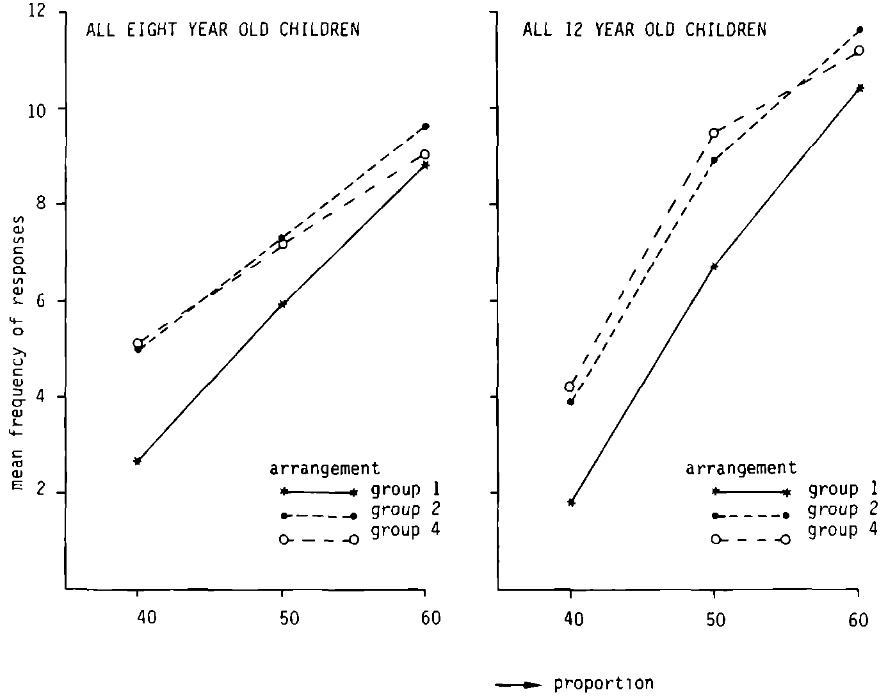


Figure 4.1. Means of responses of 8 and 12 year olds plotted against proportions for the three arrangements.

eight year old subjects appeared to be comparable with 12 year olds except that the 12 year olds appeared to be more accurate. At the age of eight, the mean frequency scores for a proportion of .40 and .50 were respectively $\bar{x}=4.5$ and $\bar{x}=6.9$, and at the age of twelve $\bar{x}=3.3$ and $\bar{x}=8.4$.

Unexpectedly, significant differences between trained and untrained eight year old subjects were not found in the multivariate testing procedure, F multivariate $(9,14)=.36$, n.s., or the univariate tests. However, the multivariate analysis on the contrasts across both groups of eight year olds revealed significance, F multivariate $(9,14)=155.15$, $p < .0001$. In the univariate test the mean score for group 2 differed from that for group 1, $F(1,22)=8.87$, $p < .01$, indicating a higher score for group 2 than for group 1 proportions, independent of the value of a proportion. A comparable effect for group 4 and group 1 proportions was not found. However, the interaction between the contrasts for group 4 versus group 1, and for .40 versus .50 was significant, $F(1,22)=$

7.40, $p < .01$, indicating that the group 4 arrangement effect was dependent on proportion (see Figure 4.1). Univariate testing further revealed that the proportion of .60 was estimated to be higher more frequently than the proportion of .50 and the proportion of .50 more frequently than the proportion of .40 ($F(1,22)=59.36$, $p < .0001$, and $F(1,22)=93.80$, $p < .0001$). Thus, hypothesis 2 needed to be rejected on the basis of the foregoing results, although the trained eight year olds showed the arrangement effect. It had to be rejected because the untrained eight year olds showed this effect too.

The effect of training on latencies of eight year old subjects

It has been predicted that latencies of eight year olds would increase by training. A multivariate comparison between trained and untrained subjects yielded non-significant results, F multivariate $(9,14)=1.94$, n.s. (see Table 4.3). Univariate testing, however, showed the predicted significant difference in grand mean of latency, $F(1,22)=15.95$, $p < .001$ (see Table 4.4). The mean for the trained eight year olds was about 2000 msec. higher than for the untrained ones ($\bar{x}=5497$ msec. and $\bar{x}=3304$ msec.). Thus, hypothesis 3 was supported by the results. Furthermore, a multivariate comparison between trained eight year olds and all 12 year olds was significant, F multivariate $(9,36)=3.42$, $p < .01$, but univariate testing did not show a significant difference in grand mean, $F(1,44)=.31$, n.s. (see Table 4.4). Therefore it may be concluded also that latencies of eight year old subjects, were close to the level of 12 year olds as a result of training.

Table 4.3.: Summary of five separate multivariate analyses of variance of latencies using contrasts as the dependent variables

Analysis	F MULTIVAR	*df*	*p*
Trained vs untrained 12 year olds	.48	9,14	n.s.
All 12 year olds	41.94	9,14	.0001
All 12 year olds vs trained 8 year olds	3.42	9,36	.01
Trained vs untrained 8 year olds	1.94	9,14	n.s.
All 8 year olds	31.26	9,14	.0001

Table 4.4.: Univariate orthogonal contrasts among means for latencies

Contrast	Comparisons								
	all eight year olds[a]			all 12 year olds[a]			trained eight year olds vs all 12 year olds[b]		
	MS	F	p	MS	F	p	MS	F	p
Grand mean	4182777080	256.84	.0001	7177834496	431.99	.0001	5156406	.31	n.s.
Proportion (.40 vs .50)	49905	.12	n.s.	3497463	3.66	n.s.	568869	.82	n.s.
Proportion (.50 vs .60)	683432	5.17	.03	16470492	85.26	.0001	2791716	17.16	.0002
Arrangement (group 1 vs group 2)	1038731	3.18	n.s.	1987847	4.18	.05	101080	.25	n.s.
Arrangement (group 1 vs group 4)	28918640	37.40	.0001	79602976	56.23	.0001	1964854	1.79	n.s.
Prop (.40 vs .50) x Arr (1 vs 2)	8073	.07	n.s.	2219053	8.17	.009	1138146	5.86	.02
Prop (.40 vs .50) x Arr (1 vs 4)	168087	1.03	n.s.	4817722	6.92	.01	1325346	3.08	n.s.
Prop (.50 vs .60) x Arr (1 vs 2)	448	.00	n.s.	16522	.06	n.s.	17156	.07	n.s.
Prop (.50 vs .60) x Arr (1 vs 4)	47288	.24	n.s.	1042397	3.80	n.s.	162693	.69	n.s.

[a] error degrees of freedom were 22

[b] error degrees of freedom were 44

The effects of arrangement on latency in eight and 12 year olds

The results of the multivariate and univariate comparisons of trained and untrained eight year olds suggested that the groups may be pooled, because of lack of significant difference. A multivariate analysis for all eight year olds using orthogonal contrasts as the dependent variables suggested significant effects, F multivariate $(9,14)=31.26$, $p < .0001$ (see Table 4.3). Univariate testing yielded a significant difference in latency between group 1 and group 4 conditions, F univariate $(1,22)=37.40$, $p < .0001$. The mean latencies for the three types of arrangements with the three proportion values are represented in Figure 4.2. In this figure it is shown that as with group 4, the mean scores for group 2 were also lower than those for group 1. This contrast, however, was non-significant, F univariate $(1,22)=3.18$, n.s.

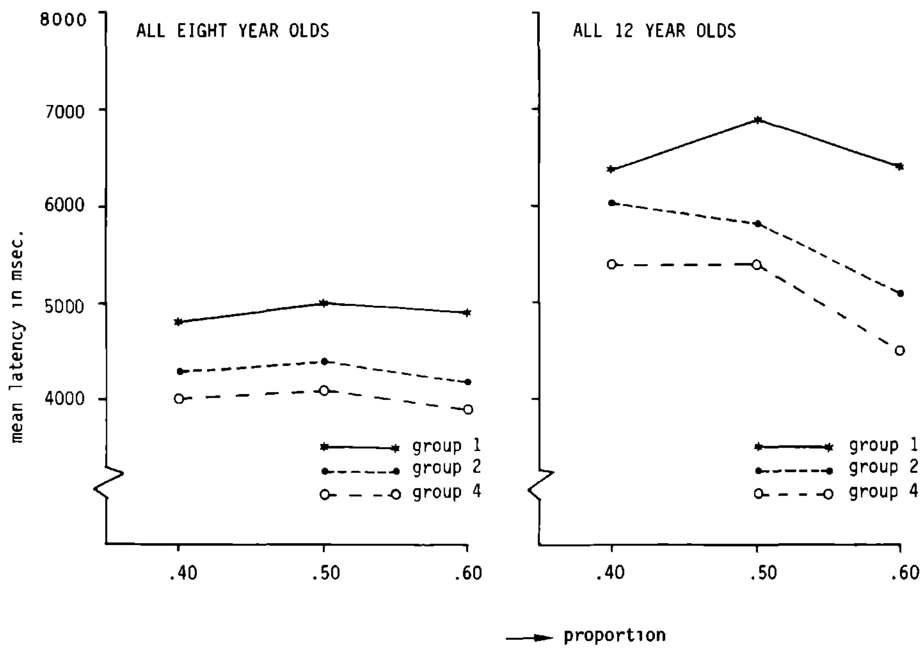


Figure 4.2.. Mean latencies of 8 and 12 year olds plotted against proportions for the three arrangements.

For testing repeated measures contrasts at the age of 12, also the two groups of 12 year old subjects could be pooled. A multivariate comparison between trained and untrained 12 year olds also was non-significant, F multivariate $(9,14)=.48$. For all 12 year olds some repeated measures contrasts appeared to be significant, F multivariate $(9,14)=41.94$, $p < .0001$ (see Table 4.3). Univariate testing yielded a significant difference in latency between group 1 and group 4 conditions, F univariate $(1,22)=56.23$, $p < .0001$,and also between group 1 and group 2, F univariate $(1,22)=4.18$, $p < .05$. The pattern of significant and non-significant interactions between arrangement and proportion at the age of 12 seems to be rather remarkable. Figure 4.2 suggests that from the .40 to the .50 proportion mean latency for grouped conditions decreased or remained equal, while it increased for group 1. On the other hand, from the .50 to the .60 proportion mean latency decreased for all arrangement conditions. Twelve year olds seemed to differ from eight year olds, or at least from trained eight year olds, with respect of this pattern of interactions. That is, in eight year olds arrangement did not interact significantly with proportion.

4.1.4. Discussion and conclusions

The arrangement effect found in the first experiment was replicated in this study. Generally, group 2 and group 4 proportions were estimated to be higher more frequently than comparable group 1 proportions. Furthermore, the proportions of .40, .50 and .60 were discriminated from each other. This was also found in the first study. Moreover, the arrangement effect appeared to be independent of training as expected for the older (hypothesis 1), but unexpectedly also for the younger subjects (hypothesis 2). However, hypothesis 2 was rejected not because of the absence of the predicted arrangement effect in trained eight year olds, but because of the presence of it in untrained subjects of the same age.

The finding that arrangement affected estimation independent of training for 12 as well as for eight year old subjects may be considered as evidence for favouring sampling by groups. Otherwise, one might argue that the absence of differences between trained and untrained groups of subjects indicates that the estimation principles presented during training were ignored by those in the trained groups and were also not applied by untrained subjects.

This suggestion, however, seems to be hardly reconcilable with the finding that training produced an increment in latency at the age of eight, supporting hypothesis 3. However, this increment may not mean that trained subjects were really doing what they were taught. It may indicate rather, that there was some conflict in a subject between performing according to the experimenter's or his or her own principles. If training produced a conflict, some such indication of it should have been found at the age of 12, because the findings suggest the same type of process for eight and 12 year olds. Furthermore, conflicting tendencies were not discovered during training. For example, subjects seemed to accept the given directions without any difficulty. Therefore, the present findings and those of the foregoing study can be interpreted as suggesting that the estimation process consisted of sampling by groups. That is, differences between trained and untrained subjects could not be found because training added nothing new to the subject's strategy.

In light of the foregoing interpretation, it seems puzzling why at the age of eight only latencies were affected by training. As a consequence of the longer latencies a higher accuracy might be expected. Such a relation between latency and accuracy has been shown in the first study for subjects who were assumed to estimate in the same way as these subjects. A possible explanation is that trained eight year olds were making more comparisons than the untrained without being able to deal effectively with all the acquired information. In fact, during training the instruction was given to make four or five comparisons. Some trained eight year olds complained afterwards that they could not remember all the outcomes. Presumably the cognitive effort required for executing the estimation process as instructed was high, especially for young subjects. Not only did numbers have to be identified and comparisons made, but also the outcomes of comparisons and four or five different locations had to be remembered. Finally an estimation had to be derived. A comparable explanation was given by Ginsburg and Rapoport (1967) for differences between about seven and eleven year old subjects in a more difficult condition of a sequential proportion estimation task.

Sampling by groups may be less demanding cognitively for older than for younger subjects. First, the latencies of 12 year olds were comparable in duration to those of trained eight year olds. However, accuracy appeared to be somewhat higher for the older subjects, indicating a more effective use of time. Second, the mean latencies of the older subjects for the different conditions compared with those of the younger subjects seem to indicate a higher flexibility in estimating. It may be assumed that in the group 1 condition relatively representative samples were drawn. Given representative sample information it would be more difficult to decide whether there are more circles or squares for a proportion of .50 than for a proportion of .40 or .60. Therefore in the group 1 condition, longer latencies may be expected, for a proportion of .50 than for a proportion of .40 or .60. Arranging circles or squares in groups of two or groups of four appeared to induce a systematic bias favouring the grouped type of figures at the expense of the ungrouped. By this bias, possibly the most ambiguous sample information was acquired at a proportion of .40 rather than .50. Therefore, an increase in latency from .40 to .50 would not be expected for grouped conditions. At the age of 12, the pattern of mean latencies for grouped and ungrouped conditions seems to be closely in line with the suggestion just made. In the group 1 condition latency increased from .40 to .50 and decreased from .50 to .60. In the group 2 or 4 conditions latency did not change or decreased from .40 to .50 and decreased from .50 to .60. For subjects at the age of eight, however, arrangement did not interact with proportion suggesting a relatively more rigid way of estimating, less adapted to the diagnostic value of the acquired sample information than for subjects at the age of 12.

It is suggested in this discussion that younger subjects estimated more rigidly, because application of sampling by groups required more cognitive effort to them than to older subjects. However, another explanation is also possible. At the age of 12 a cognitive structure for interpreting sample information statistically may have been developed based on the concept of probability (e.g. Fischbein, 1975, and Piaget & Inhelder, 1951). This concept may have been developed at that age, and therefore a possibility exists that the 12 year old children made use of it. Although the present data were not designed to test for the best of these two alternative explanations, some critical remarks with respect to the latter position can be made. First, the systematic biased responding induced by arrangement with eight and 12 year olds and adults suggests the absence of statistical thinking in acquiring sample information, controlling for representativeness of it. If there is no evidence of such a way of thinking in acquiring it, why should statistical principles be applied in deciding on acquired information? An explanation that does not presuppose such a structure seems more parsimonious. Second, perhaps with age there may be more reflection on acquired information. The present and previous study suggest that skill in applying sampling by groups may be considered as very relevant in this respect. With age subjects seem to be more able to apply sampling by groups in an integrated and refined way.

4.2. Experiment III: an examination of age differences in the use of sampling by groups

4.2.1. Introduction

The first study suggested that subjects ranging in age from about 12 years to adulthood estimated relative quantities comparatively, that is, by sampling by groups. Some of the findings in the second study, however, suggested the same type of process for children of about eight years old. However, although sampling by groups has been attributed to the age range from eight years to adulthood, the complete absence of age differences cannot be assumed. On the contrary, differences in the application of this process are suggested, and the present study was directed at these differences.

It has been hypothesized that in sampling by groups numerical comparisons are made iteratively. Some findings seem to indicate that the iterative aspect of this process becomes cognitively less demanding with age. Older subjects seem to be more capable of screening a stimulus configuration effectively for a relative longer period of time than younger. In the second study, for example, latencies of eight year olds appeared to be increased by training, however without any positive effect on accuracy. In the first study on the other hand, adults used more time for responding than 12 year olds, and estimated relatively more accurately with less bias. Furthermore, 12 year olds used more time than untrained eight year olds, and also estimated more accurately than untrained or trained eight year olds. In addition, the better performance of adults in the longest exposure time condition could be attributed to their relatively longer latency. These findings seem to suggest that estimating by older subjects would be comparable to that of younger subjects if their latencies were shorter, but not that younger subjects would perform as accurately as older ones if the stimulus material were presented for a longer period. That accuracy decreases and biased responding may increase with shorter latencies has been shown in the first study.

Another piece of evidence in line with the hypothesis that the iterative aspect of estimating requires less cognitive effort with older subjects is the presumably increased adaptability of these subjects to variation in ambiguity of sample information. This change with age was suggested in the second study by the patterns of latencies of older and younger subjects. Twelve year olds seemed to take longer to respond when they were acquiring sample information that could be considered relatively more ambiguous. Such a differentiation in latency was not found in eight year olds. Therefore, it was suggested that with age estimating becomes more flexible.

With these empirical arguments there are also some rational ones which emphasize the developmental interest of iteration. The more numerical comparisons that are made, the greater the load put on memory. For each subsequent comparison, not only does prior sample information have to be retained, but also the location of a stimulus configuration in the array from which it was extracted, assuming that a subject tries to inspect as much as possible of a stimulus configuration. It is a widely accepted point of view that performance in memory tasks increases with age (see e.g. Chi, 1977; Flavell, 1977; Huttenlocher & Burke, 1976). In addition, the importance of memory for explaining age differences in non-memory tasks has been shown in some other studies. For example, Beckwith & Restle (1967) demonstrated the function of memory with respect to age for enumeration. Besides, the importance of memory for the development of transitive reasoning has been shown by e.g. Bryant (1974), and Trabasso, Riley & Wilson (1975).

The contribution of iteration to age differences in the application of sampling by groups can be demonstrated by reducing the opportunity for it. The number of numerical comparisons that can be made may be manipulated by the exposure time of a stimulus configuration. By reducing presentation time, the opportunity of iteration should also be reduced. In fact, such a procedure was applied in the first experiment. However, in that study estimations given for the shorter presentation time might have been influenced by prior presentation of the same stimulus configurations to the same subjects for a longer time. The rather unexpected differences between adults and 12 year olds for the one second presentation time condition were interpreted in this sense. In the context of the iteration hypothesis, such an interpretation seems to be plausible. Yet a reduction of the opportunity for iteration should minimize age differences. To investigate the contribution of iteration more carefully, in the present study the stimulus configurations of the second study were shown for one second only to eight year olds, 12 year olds, and adults. The absence of age differences in estimation for this range of age with this short presentation time would be considered as evidence favouring the iteration hypothesis.

4.2.2. Method

Subjects

Twelve undergraduate psychology students at the University of Nijmegen, and 24 school children served as subjects. The children included: 12 sixth graders with a median age of 11-11 (range 11-10 to 12-7), and 12 second graders with a median age of 7-6 (range 7-0 to 8-6), all from the same elementary school in Nijmegen. Boys and girls were equally divided at each age level. The undergraduates volunteered to be subjects in the experiment without payment. However, the school children were given comics afterwards, in appreciation of their participation.

Procedure

The 108 stimulus configurations of the second experiment were presented in exactly the same way as was done in that study, except for presentation time. This was one second for each stimulus configuration. Before presentation of this series instruction was given with the instruction series, practice series, and the directions used in the first experiment. However, all stimulus configurations of the practice series were shown for only one second each.

4.2.3. Results

For comparing the age groups on the effects of proportion and arrangement, and for analyzing the effects of these variables on estimation across age groups, nine orthogonal contrasts were made in the same way as in the second experiment. Age comparisons were made for adults vs. 12 year olds, and for adults and 12 year olds vs. eight year olds (see Tables 4.5, 4.6) Proportion and arrangement effects were tested across all age groups, using the nine orthogonal contrasts (see Table 4.6).

Table 4.5. Summary of three separate multivariate analyses of variance of frequency scores using contrasts as the dependent variables

Analysis	F MULTIVAR.	df	p
12 year olds vs adults	.24	9,25	n.s.
12 year olds and adults vs 8 year olds	1.60	9,25	n.s.
Across all ages	231.94	9,25	.0001

Comparison of age groups

The multivariate comparison between adults and 12 year old

Table 4.6.: Univariate orthogonal contrasts among means for frequency scores

Contrast	Comparisons								
	12 year olds vs adults[a]			12 year olds and adults vs 8 year olds[a]			across all ages[a]		
	<u>MS</u>	<u>F</u>	<u>p</u>	<u>MS</u>	<u>F</u>	<u>p</u>	<u>MS</u>	<u>F</u>	<u>p</u>
Grand mean	18.96	0.60	n.s.	6.32	0.20	n.s.	15431.11	487.24	.0001
Proportion (.40 vs .50)	3.17	0.77	n.s.	5.57	1.36	n.s.	1146.67	279.55	.0001
Proportion (.50 vs .60)	1.17	0.31	n.s.	5.11	1.36	n.s.	298.68	79.47	.0001
Arrangement (group 1 vs group 2)	0.45	0.12	n.s.	1.11	0.30	n.s.	161.00	43.18	.0001
Arrangement (group 1 vs group 4)	1.36	0.12	n.s.	14.08	1.24	n.s.	35.04	3.08	n.s.
Prop (.40 vs .50) x Arr (1 vs 2)	1.11	0.61	n.s.	8.11	4.46	.05	5.58	3.07	n.s.
Prop (.40 vs .50) x Arr (1 vs 4)	1.84	0.65	n.s.	12.76	4.51	.05	22.69	8.01	.008
Prop (.50 vs .60) x Arr (1 vs 2)	0.17	0.14	n.s.	2.78	2.26	n.s.	1.22	1.00	n.s.
Prop (.50 vs .60) x Arr (1 vs 4)	0.09	0.04	n.s.	2.53	1.11	n.s.	1.56	0.69	n.s.

[a] error degrees of freedom were 33

subjects on the repeated measures contrasts yielded no significant effects, F multivariate (9,25)=.24, n.s. (see Table 4.5). In addition, none of the univariate tests was significant (see Table 4.6). The comparison of adults and 12 year olds vs. eight year olds was also not significant, F multivariate (9,25)=1.60, n.s. The univariate tests suggested differences with respect of the interactions between the proportions of .40 and .50 and group 1 and group 2, F univariate (1,33)=4.46, p < .05, and group 1 and 4, F univariate (1,33), p < .05. Figure 4.3 shows that the mean score for group 4 is higher than that for group 1 in all conditions, except for the proportion of .40 in adults and 12 year olds. So for at least this proportion value there is no arrangement effect in the two older groups. At the age of eight, however, there may be an arrangement effect for this proportion.

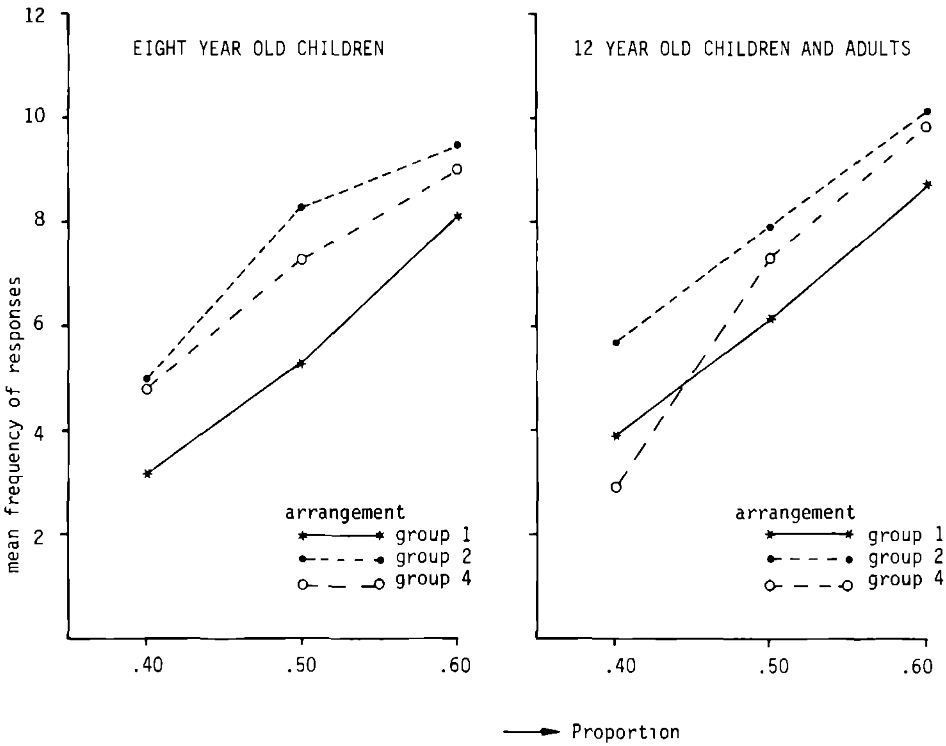


Figure 4.3. Mean of responses of 8 year olds, and 12 year olds and adults plotted against proportions for the three arrangements.

The effects of proportion and arrangement

On the basis of the age group comparisons, age groups were pooled for testing the effects of proportion and arrangement. The absence of an arrangement effect for group 4 at the proportion of .40 in adults and 12 year old children will lead only to a more conservative test of the effect of the group 4 arrangement.

The multivariate test of the nine contrasts across all age groups was highly significant, F multivariate $(9,25)=231.94$, $p < .0001$. According to the univariate testing procedure, the mean score for .40 differed from that for .50, and the mean score for .50 differed from that for .60 (F univariate $(1,33)= 297.55$, $p < .0001$, and F univariate $(1,33)=79.47$, $p < .0001$ respectively). Figure 4.3 shows that the mean scores for the three proportions were ordered as follows: $(.40 < .50 < .60)$. Moreover, group 2 arrangements differed from group 1 arrangements F univariate $(1,33) =43.18$, $p < .0001$, and were estimated to be higher more frequently. A comparable effect was not found for group 4, F univariate $(1,33)=3.09$, n.s. However, a significant interaction between groups 1 and 4 at the proportions of .40 and .50, F univariate $(1,33)=8.01$, $p < .01$, was obtained.

4.2.4. Discussion and conclusions

The findings of the present study strongly support the hypothesis that iterative sampling requires more cognitive effort at younger than at older ages. There were no or small age differences in estimation for the range of eight years to adulthood when the configurations were presented with a short exposure time greatly reducing the opportunity for iteration. These results were clearly at odds with some findings of the one-second exposure time condition of the first experiment. In that condition adults estimated more accurately and with less bias than 12 year olds. However, an explanation of these results was that repeated presentations of the same patterns for longer exposure periods to every subject differentially affected the two age groups. The present results may be regarded as supporting that explanation indirectly.

For a better understanding of the cognitive effort that iterative sampling requires at different ages more precise definition and explanation are needed than were available in this study. All processes related to identifying small groups of objects in an array and comparing numbers more than once may be regarded as iteration factors. In view of the evidence, it would appear that the differences between the estimation performance of eight year olds, 12 year olds, and adults is due to more than just speed of sampling. In the first study it was suggested that adults, in comparison with 12 year olds, use their longer response time to screen a configuration more carefully, and not just for making more numerical comparisons. It may be that younger children are less able to remember material identified earlier such as locations or outcomes, or are less selective and strategic in the ways they screen a configuration than older children or adults. Furthermore, each group identification may be a source of age differences. When younger children have more difficulties with group identification than older children or adults, one may expect that they will have even more difficulties with repeated identication, because previously acquired information must also be remembered. A better understanding of the development of the way in which sampling by groups is applied requires a careful examination of either of these aspects separately and in combination with each other. Measuring of scanning patterns at different ages could be very helpful in this respect.

4.3. Experiment IV: a second training study

4.3.1. Introduction

In the previous studies, using sampling by groups as method of estimation has been inferred from the presence or absence of the arrangement effect. This effect has now been shown several times using the same task and stimulus materials with various procedures. It has been described as a systematically biased way of estimation consisting of a general tendency to estimate grouped proportions to be as more frequent than comparable ungrouped ones. Some specific empirical evidence, however, for inferring sampling by groups from the arrangement effect has not yet been given. A convincing argument could be made if this effect could be induced by teaching the inferred process to children who are not expected to show it. In the previous training study children of about eight years old were trained to show the arrangement effect. However, in that study untrained eight year olds also showed the effect and thus the dependency of it on training could not be established for those children. Therefore, in this study younger subjects, approximately six year olds, were trained.

A pilot training study using children of about six years old and results of the previous studies suggested a few modifications of the design of the present training study. These specific changes are summarized in the following paragraph.

First, children were instructed to make one comparison rather than a few. This was done because of the increased effort required for making iterative comparisons at younger ages. Furthermore, the arrangement effect, if it exists, should occur with only one comparison. Second, the results of the pilot study suggested that the practice given in the previous training study should be expanded, especially for making decisions on the basis of a perceived number of circles and squares. In the pilot study training did not affect estimation, although children said that they used the procedure. However, doubt about the validity of these verbal reports seem justified (see e.g. Nisbett & De Camp Wilson, 1977). In regard to non-application of newly acquired strategies, Flavell (1977, p. 199) suggested that an old strategy may compete with and win from a new one because the old strategy "has been in the child's repertoire longer, is less difficult and effortful to execute ..." Third, the previous studies showed a negligible difference between group 2 and group 4 conditions suggesting that the arrangement factor may be reduced from three to two levels. Fourth, the three levels of proportion may also be reduced. In this study only proportions of .40 and .50, in group 1 or 2 arrangements were presented. Thus, less time was required for the whole series of stimulus configurations, and the training could be given just before presentation of the series to maximize training effects. Finally, some children without training in the pilot study estimated grouped proportions as higher more frequently than the comparable ungrouped ones. It would be difficult to show the training effect with children whose pretraining estimations were comparable to those predicted after training. In this study the same series of stimulus configurations was presented twice.

Children with higher individual scores for the grouped conditions on the first presentation were treated separately. They were given training to explore whether they could be trained more easily than the other children, but were omitted from further analyses. These subjects were identified as the advanced group. The other children were randomly assigned to be in the control or training group. Training was given to the training group and the advanced group just before the second presentation of the series. The arrangement effect was predicted for the children in the training group.

4.3.2. Method

Subjects

Subjects were 33 kindergartners from two nursery schools, one in Nijmegen and one in the neighbourhood of Nijmegen. All children were divided into three groups: a training group (n=12) with a median age of 6-1 (range 5-9 to 6-6), a control group (n=11) with a median age of 6-1 (range 5-9 to 6-8), and an advanced group (n=10) with a median age of 6-1 (range 5-8 to 6-7). The number of boys and girls within each group was approximately equal. Each child was given a packet of feltpens, in appreciation of her or his participation.

Materials

The series of 108 stimulus configurations of the second and third experiment was reduced to a series of 48. The new series was composed exclusively of configurations representing a proportion of .40 or .50 and an arrangement of group 1 or group 2, that is, 12 configurations for each proportion by arrangement combination. The original series was divided into six blocks of 18 stimulus configurations. By removal of all configurations representing a proportion of .60 and/or an arrangement of group 4 eight configurations remained within each of these six blocks. By combining the configurations of block one and block two, three and four, five and six, respectively, the number of blocks was reduced to three of 16 configurations each.

The instruction and practice series used in the previous experiments were also modified. In the instruction series, the two configurations with a group 4 arrangement were left out. In the practice series the two group 4 configurations were replaced with two comparable group 2 arrangements. The series was completed with two additional group 1 configurations, so that the whole practice series consisted of eight configurations, two for each combination of proportion (.60 or .40) and arrangement (group 1 or group 2).

A training series of 16 stimulus configurations, increasing in total number of figures but of approximately equal density, was constructed by computer. Each configuration consisted of circles and squares. The series consisted of four blocks of four configurations each. The first block contained configurations with a total number of 20 figures, the second with a total number of 40, the third of 60, and the last of 120 figures. The proportion of circles or squares was .40 or .60 arranged according to group 1 or group 2. Thus, contrary to the main series, configurations with a proportion of .50 were omitted. This was done in order to avoid any suggestion to the child during the training that some configurations might exist of the same number of squares and circles. Within each block all four arrangement by proportion conditions were represented by a configuration. Within a block, circles as well as squares were placed grouped at a proportion of .60 *or* .40. However, within two subsequent blocks they were placed grouped at a proportion of .60 *and* .40. Each configuration of the training series was printed on a separate piece of paper 20.9 cm. x 29.6 cm. (see Appendix 2). In addition to the training series, seven different examples of samples were made, each consisting of a small number of circles and squares (see Appendix 2). Within an example the number of circles and squares or vice versa could be 2,3; 2,4; and 3,4 respectively. An example was a part of a stimulus configuration belonging to the group 2 or group 1 conditions. They were printed separately on one piece of paper of 20.9 cm. x 29.6 cm., each surrounded by a rectangular frame of 6.1 cm. x 5 cm. Finally, two additional configurations were constructed by computer, each consisting of a total of 120 circles and squares and printed on a separate piece of paper of 20.9 cm. x 29.6 cm. one configuration represented one of the group 1 conditions (.60 squares), the other one of the group 2 conditions (.40 squares arranged in groups).

Procedures

The series of 48 stimulus configurations was presented twice to each subject with an interval of one week between sessions. For each session the order of blocks was counterbalanced for each subject. At the beginning of the first session the instruction procedure used in the previous experiments was administered using the modified instruction and practice series. After that the main series was presented. The configurations of the practice and main series were presented for seven seconds each in the same way as was done in the previous experiments.

After the first presentation of the main series children were assigned to three groups (training, control, advanced). Figure 4.4 shows the pattern of scores for inclusion in the advanced group. Children who estimated group 2 arrangements to be higher more frequently for at least one proportion (.40 or .50), and equal or higher for the other were placed in the advanced group. Children whose individual scores did not meet this criterion were randomly assigned to the control or training group.

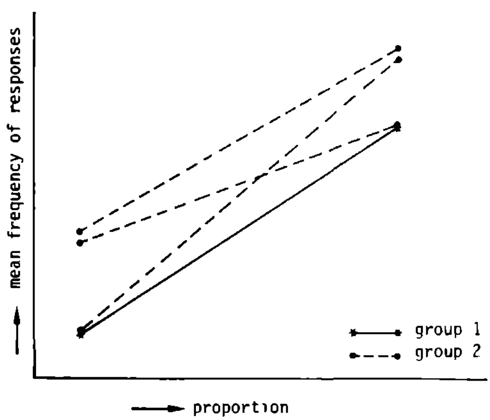


Figure 4.4.: Schematic representation of the pattern of scores for inclusion in the advanced group.

The second session started with a training period of about 30 minutes for the children in the advanced and training groups. However, the children in the control group were given easy puzzles to solve for about the same amount of time at the beginning of the session. After that, the practice series and then the main series were presented again. Children in the training group were told to apply the procedure they were taught. Children in the control group were reminded of the task and instructions of the previous session. Apparatus, and other procedures were the same as in the first session, except for the distance from the T.V. monitor to the place where the child was seated. The monitor was placed at such a distance from the child that he or she could easily touch the screen with the tip of his or her fingers. Doing this, the trained procedure could be applied more easily by children in the training group. To make the situations comparable, the same was done for the control group children.

At the beginning of the training session the child was asked whether she or he remembered the pictures and tasks of the previous session.* Every child answered this question affirmatively. Then the child was told that she or he would be taught how adults and older children judge whether there are more circles or squares. After this the experimenter showed the seven examples to the child. The child was told that these examples were samples, and each consisted of a small number of circles *and* squares so that the number could be seen immediately. After that the child was asked how many circles and squares there were in each sample, and whether there were more circles or more squares. All children answered these questions correctly without apparent difficulty. After that, the child was told that each of the seven samples came from a larger 'picture' while the experimenter showed the two additional configurations. Further, the child was told that a sample could also be made in another way. A white cardboard of 20.9 cm. x 29.6 cm. with a rectangular hole of 6.1 cm. x 5 cm. in the midst of it was shown to the child and placed on one of the additional configurations such that figures could not be seen except through the hole. Then, the child was asked to draw a few samples from the two additional stimulus configurations using the cardboard. After

---

*) The specific wording for the instructions in Dutch may be obtained by writing to the author, Dept. of Devel. Psychol., Psychol. Lab., University of Nijmegen.

that, the first stimulus configuration of the training series was presented, and the experimenter explained the use of sampling. The child was told that within a sample one could see immediately and correctly whether there were more circles or more squares in the whole configuration. The plausibility of this rule was demonstrated to the child using the first block of the training series in the following way. For each configuration, first a sample was drawn by the experimenter using the cardboard. Then, the child was asked to name what he or she perceived to be the *number* of circles and squares in the sample, and whether there were more circles or squares in the sample and in the whole configuration. It was emphasized by the experimenter that the last two answers should be the same. After an estimation had been given, the child verified the correctness of it by counting all circles and squares of the stimulus configuration. Estimations were only confirmed by counting on the first presentation of the first block of the training series. After presentation of the first block the child was instructed to do the sample drawing by him or herself, to estimate on the basis of just *one* sample using the taught procedure, and to omit counting afterwards. Then the remaining blocks were presented.

Presentation of the whole series was repeated twice. On the second and third presentation, the child used his or her *fingers for sampling* by pointing somewhere in the midst of a configuration. Moreover the estimation procedure was slightly shortened. Instead of naming the most frequently occurring type of figure twice, once for the sample and once for the whole configuration, the child's estimation was provided immediately after the number of circles and squares in the sample were given. The experimenter ensured that each estimation was compatible with the numbers named by the child. On the third presentation, the configurations were shown on the video monitors, controlled by computer. The monitor was placed close enough so that the child could touch its screen easily with his or her fingertips. Besides, the series was divided into two halves. From each block configurations representing the group 1 and the group 2 conditions were randomly placed in the first half and the remainder in the second half. The blocks were assigned to each half in order of the number of figures in a configuration. Thus configurations with a smaller number of figures were presented first. The first half was presented untimed, the second timed, that is, at a constant value of seven seconds for each configuration. The estimation procedure was unchanged except for number naming. Naming the number of circles and squares in a sample was gradually omitted during presentation of the second half. In the second half, the whole half was repeated when a child was not able to fluently provide an estimation by the last few configurations, that is, give an answer before the next presentation. Only a few children needed such a repetition of the second half. Before the second presentation of the practice and main series, children who received training were encouraged to apply the estimation strategy that had been taught and to make a sample using their fingers. In the training session each child was seen individually.

4.3.3. Results

The control and training group were compared for the effects of proportion and arrangement on frequency scores using a multivariate analysis of variance. Dependent variables were the following eight orthogonal contrasts formed on the repeated measures:

1) the multivariate grand mean across all contrasts compared to zero;
2) time of testing (pretest vs. posttest);
3) proportion (.40 vs. .50);
4) arrangement (group 1 vs. group 2);
5) time of testing x proportion;
6) time of testing x arrangement;
7) proportion x arrangement;
8) time of testing x proportion x arrangement.

A significant difference between the training and control group may be expected especially with respect to the interaction between time of testing and arrangement on frequency scores.

The multivariate comparison between the control and training group on the repeated measures contrasts yielded significant effects, F multivariate $(8,14)=5.45$, $p < .003$ (see Table 4.7). The univariate testing procedure showed that the control and training group differed from each other in the contrasts time of

Table 4.7.: Univariate orthogonal contrasts among means for frequency scores

Contrast	Comparison					
	training group vs control group[a]			training group posttest[b]		
	MS	F	p	MS	F^c	p
Grand mean	87.04	7.30	.01	1788.52	243.70	.0001
Time (pretest vs posttest)	48.02	6.00	.02			
Proportion (.40 vs .50)	5.40	2.12	n.s.	3.52	1.70	n.s.
Arrangement (group 1 vs group 2)	53.88	6.70	.02	150.52	26.39	.0004
Time x proportion	3.41	1.33	.004			
Time x arrangement	57.52	10.26	.004			
Proportion x arrangement	7.68	3.12	n.s.	7.52	3.10	n.s.
Time x prop x arr	3.66	1.50	n.s.			

[a] error degrees of freedom 21, F multivar. $(8.14)=5.45$, $p < .003$

[b] error degrees of freedom 11, F multivar. $(4,8)=56.29$, $p < .0001$

[c] the number of orthogonal contrasts in the numerator was four.

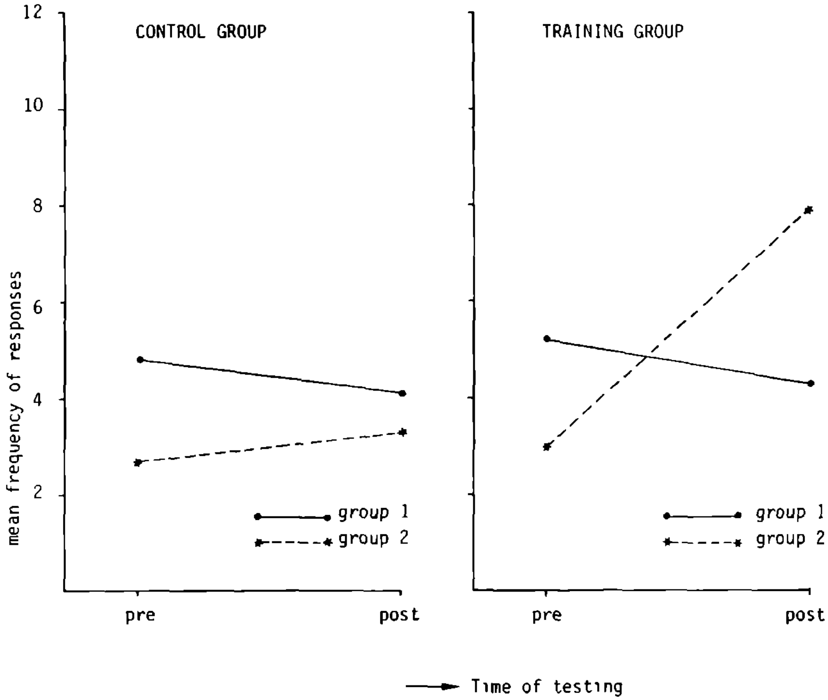


Figure 4.5.: Interaction between group, time of testing and arrangement.

testing, arrangement, and arrangement x time of testing (respectively, F univariate $(1,21)=6.00$, $p < .02$, F univariate $(1,21)=6.70$, $p < .02$, and F univariate $(1,21)=10.26$, $p < .004$. Group 2 arrangements were estimated to be higher more frequently as a consequence of training as can be seen in Figure 4.5.

A separate multivariate analysis of variance was performed on the posttest scores of the training group, using four orthogonal contrasts as the dependent variables. The following contrasts were formed on the repeated measures: (1) the multivariate grand mean across all contrasts compared to zero, (2) proportion, (3) arrangement, and (4) proportion x arrangement. The multivariate test of the four contrasts was highly significant, F multivariate $(4,8)=56.29$, $p < .0001$ (see Table 4.7). The univariate testing procedure yield a significant arrangement effect F univariate $(1,11)=26.39$, $p < .0004$. Group 2 proportions were estimated to be higher more frequently than the comparable group 1 independent of the value of proportions.

4.3.4. Discussion and conclusions

Two conclusions seem to be warranted from the present study. First, the arrangement effect can be induced by teaching sampling by groups to children who did not show it prior to training. So the hypothesis that the arrangement effect is based on sampling by groups is supported. Second, children at the age of about six are able to learn to sample by groups, that is to estimate on the basis of a small sample. One might argue, however, that both the conclusions are based on group means, but claims that a strategy is used by individuals should be based on individual data. Without discussing the merits of this argument, it can be stated with some certainty that the foregoing conclusions seem justified with respect to the individual data. After training, the individual scores of all 12 training group children and of none of the control group children met the criterion used for selecting the advanced group.

The individual pretest score of some children, identified as the advanced group, suggested that sampling by groups was applied by them without training. These children were omitted from analyses because obviously the predicted training effect could not be demonstrated with them. However, they were given training to explore whether they could be trained more easily than the children in the training group. Although trainability was not measured as such, the subjective impression of the trainer was that training was easier with them. Furthermore, highly significant arrangement and proportion effects were found for this group independent of training, or time of testing (F univariate $(1,9)=65.39$, $p < .0001$, and F univariate $(1,9)=34.86$, $p < .0003$). The proportion effect may be considered rather remarkable, because such an effect was not found with the training group, and may hardly be expected on the basis of the training given. Although only one sample was made for each estimation, sampling by the advanced group children was probably more in line with the real proportion values than was the case with the training group children, thus suggesting more advanced estimating by the advanced group.

A basic developmental question is why sampling by groups is not applied spontaneously by most six year old children. The answer can hardly be that most of these children do not have the are several reasons for saying this. Among them, the following seems to be of most direct relevance. First, training appeared to be successful, although number conservation, for example, was not trained. Furthermore, at the age of about six most children show the ability to conserve numbers (see e.g. Brainerd, 1978; Flavell 1977). Besides, if not estimating by groups at the age of six is simply a question of set, training should have affected estimating in the pilot study also. Finally, in the present study as well as in the pilot study, children seemed ready and able to accept the directions. However, application of the procedure, especially identification of groups seemed to be very difficult to them. Therefore, in the present study, children were forced to do so by naming the numbers of a sample. Moreover, they were instructed to use their fingers for identification during the presentations. Perhaps sampling by groups requires too much cognitive effort of six year old children. Therefore, a less mature strategy that may be performed with less effort than sampling by groups is used.

5. THE DEVELOPMENT OF ESTIMATION RETROSPECT AND PROSPECT

The present series of experiments was designed to develop tentative answers to the following general questions:

1) How does a person estimate the relative quantity of simultaneously presented collections, and
2) does the estimation process change with age, that is, within the range from about six years to adulthood.

In the introduction (Chapter 2), the possibility was suggested that from the age of about five to seven a method of estimation is developed that is based on numerical rather than spatial information. Furthermore,it was hypothesized that numerical information may be acquired by identification of small groups and quantification of the number of objects within a group by subitizing. This process was labelled sampling by groups. It was assumed that a configuration is inspected several times depending on its exposure time. On each inspection, a number of circles and squares is sampled, and a comparison is made between these numbers resulting in some outcome, e.g. 'more squares'. Finally, it was predicted that this process will lead to systematically biased estimations in some situations. That is, given configurations consisting of different types of arrangements for different types of figures such that small subitizable groups may be identified more easily for one than for another category, estimations will systematically favour the category with the relatively more easily identifiable groups. This prediction was confirmed repeatedly, and the arrangement effect generally appeared from the age of eight. However, it could also be induced at a younger age by teaching sampling by groups, and children of about six years old who did not show the effect prior to training did demonstrate it afterwards. So the finding of the present studies suggest that, at least from the age of eight, numerical information is used for estimating whether there are more objects of one or another collection and furthermore, that this information may be acquired by sampling groups.

Other questions may be raised on the basis of the present findings. One concerns the development of sampling by groups, a second the generality of the present findings across different types of tasks, arrangements, and presentation methods, and a third, the component processes of sampling by groups. In this final chapter attention will be directed primarily to the developmental aspects, because these were the main concern of the present studies. The generality problem will be discussed only incidentally within the context of the discussion about the development of sampling by groups. Finally, the chapter will conclude with some suggestions for future research with respect to the third question, especially in regard to the control and monitoring of the sampling activity.

The first developmental question concerns what determines whether young children will or will not estimate on a numerical basis.

Some findings of the last study suggest that at least for children of about six years old the answer may not be given in terms of lack of number knowledge or set to apply the available principles because, as discussed in chapter 2.1, at that age most children show the ability to conserve number. Furthermore, they are able to count or subitize, and presumably consider numerical information to be relevant to estimating. Moreover, Gelman (1978) and Gelman & Gallistel (1978) have shown recently that number knowledge of young children may have been considerably underestimated by researchers. They have presented strong arguments in favour of the position that at the age of 2½ children already know how to count, but commonly seem not to apply their knowledge in a skillful and culturally acceptable way. So, the available knowledge may be sufficient to compare quantities on a numerical basis. On the other hand, the acquisition of numerical information seems to be very hard for six year olds and younger children. One may infer that, at that age, most children are not able to identify and subitize small groups within a larger array of figures without the additional support of pointing with their fingers. Thus, even if the children wish to estimate numerically, they are probably not able to do so without this kind of support.

Apart from the help of pointing, some aid was also available from the way one type of figure was arranged in most stimulus configurations. Figures were in separate groups of two or four for one type of figure. However, even with this help the younger subjects needed to point. Therefore, part of the answer to the question of what precipitates estimating on a numerical basis is the amount of support for estimating in that way which exists in the task. Thus, if the configuration may be easily arranged in small groups, and other needed support such as pointing is possible, even young children will appear to use this method. With other kinds of estimation tasks it is not surprising that other cues were found to be important. For example, length has been found to be important for estimating from linear arrays (see Chapter 2.1).

Another developmental question concerns possible changes in estimating with age, or, what aspect of estimation is developing with the child.

From the present findings one might infer that between the age of six and eight the way of estimating changes fundamentally. As compared to eight year olds, most children of about six commonly do not estimate on the basis of number. Presumably, they are attending to spatial characteristics, perhaps visual area, rather than number. However, from the point of view presented in the last paragraph it follows that this difference may be task specific. Estimating non-numerically was interpreted as a problem of information acquisition rather than missing principles. In line with this position, attending to numbers by eight year olds may be a question of increased skill in abstracting numbers from a configuration rather than knowledge of estimation principles. In the present studies, it appeared that eight year olds estimated using sampling by groups, while most six year olds used a simpler strategy, probably based on the visual area of the two types of figures in a configuration. From the age of eight on, sampling by groups was apparently used at all ages. However, older persons seemed to be more skillful in doing this than younger ones. So, what may be developing with age is an increase in skill of sampling by groups.

The greater skill may be demonstrated by an increase with age in use of sampling by groups across different task situations. Although this hypothesis has not been investigated directly by the present studies, it is in line with the interpretation that has been given of the performance differences between children of about six years old and older persons. Besides, the higher skillfulness may also be demonstrated by an increase with age in the flexibility and efficiency with which numerical information may be abstracted from an environment and integrated for a given type of tasks. For example, in the present studies older persons estimated more accurately than younger ones. However, their higher accuracy could be attributed to their longer response times and disappeared when configurations were presented with a relatively shorter time of one second. Moreover, it was suggested by the second study that older subjects were more adapted to variation in the ambiguity of acquired sample information.

The point of view presented in this chapter about the development of sampling by groups is closely in line with a recent position taken by Gelman & Gallistel (1978) about the development of counting. According to these authors, from the age of 2½, children primarily learn to apply their available knowledge about counting. Furthermore, the view of increasing skillfulness fits with some of the ideas of Schaeffer (1973), Schaeffer et al (1974), about cognitive development and Bruner (1973) about sensori-motor development in infants. According to these authors, a mechanism, such as described by Schaeffer (1973) as automation and by Bruner (1973) as modularization, is essential to the development of cognition. Automation reduces the working memory capacity required for processing of a strategy or program. Thus, the possibility is created to direct attention to new aspects of a task during processing.

A third developmental question concerns possible determinants of development.
The initiative or motivation for development is an important topic of concern to developmental psychologists. Although the present studies were not directly designed to study this question, it will

be discussed here because studying estimation at different ages may be of specific concern to thinking about cognitive development. It will be shown that some widely accepted positions do not hold for explaining the development of estimation.

One widely accepted Piagetian point of view is that development is induced by cognitive conflicts (Piaget, 1971). According to this view, cognitive conflicts lead to a state of disequilibrium in the organism, and a tendency to look for new higher-order principles which restore equilibrium. Cognitive conflict may be indiced by presenting negative feedback or by showing the individual that application of two or more principles, he or she believes in, leads to conflicting answers (see e.g. Heymans, 1979). For instance, in one of their training studies, Inhelder, Sinclair & Bovet (1974) showed children that roads to be compared sometimes were equal in length, and sometimes not, depending on the principle used. Children compared roads on the basis of correspondence of end points or the number of separate sections into which each road was divided. To induce a cognitive conflict, roads were arranged so that using both principles led to different answers. In this study it was suggested that children would develop the principle of measurement, or division into standard units, to solve this and comparable conflicts.

Recently, the cognitive conflict principle has been elaborated as a consistency checking mechanism, by Klahr & Wallace (1976). According to this view responses given by a subject to the same type of problems are screened for consistency. In case of inconsistency among answers, the principles used for a given category of problems are rejected. After that, new principles are developed and checked as candidates to solve the problem, and only principles that generate consistency are accepted. Without extending discussion of the cognitive conflict and comparable principles any further here, they seem unlikely for application to the development of estimating. A cognitive conflict principle may operate for algorithmic type of problems for which correct responses are guaranteed by application of the right principle to every problem of the same type. Without this guarantee the criterion of consistency does not function. The quality of the estimation principle used cannot be judged on the basis of the correctness of each response, but rather on the degree of accuracy in the long run. It may be doubted seriously whether human beings are able to do this (see e.g. Kahneman & Tversky, 1972; Wagenaar, 1972). This suggests that the development of principles of estimating must be explained in another way. With respect to skill development, practice and experience were suggested by Gelman & Gallistel (1978) as possible sources. Perhaps, the development of estimation principles can be attributed to comparable determinants. Finally, a suggestion made by Schaeffer et al (1974) seems to be of relevance in this respect: "... number development is determined more by application of number skills to object arrays than by thoughts about numbers in the absence of arrays or by spontaneous cognitive reorganizations, i.e., equilibrium" (p. 378).

In summary, according to the position presented in the previous paragraphs, the processing of numerical information is treated as a basic, and with age increasingly general, characteristic of estimating relative quantities. Furthermore, it is assumed that subitizing functions as the main quantification mechanism in estimating the relative quantity of simultaneously presented collections.

The point of view presented here, should be investigated further with other types of stimulus materials and tasks. Furthermore, how groups are identified and which features are used for identification, the number of groups identified and quantified within one fixation or inspection, and the number of elements within a group given different types of arrangements are suitable areas for research. By examining these and comparable problem areas, increased psychological understanding may be acquired of situations and factors that lead to systematically biased ways of estimating, such as the arrangement effect shown in the present studies, or the solitaire illusion (Frith & Frith, 1972).

Other questions are the ways accumulated numerical information is integrated into estimation and the effects of previously acquired information on the exploration of a pattern.

The latter questions specifically concern the control of the sampling activity, or, the mechanisms which determine where in a configuration sampling is started, how it is continued and when and where it is stopped. The increase in skill of estimation with age, as suggested by the present studies, does presuppose the development of a more general plan or program by which iterative sampling is controlled. Especially the results of the third study indicating that age differences in estimating can be minimized by reducing the opportunity of iterative sampling are of relevance in this respect. Furthermore, comparable suggestions were made e.g. by Brown & DeLoache (1978).

A program which controls iterative sampling may be tuned to data of different sources. Possible candidates, for example, are: the number of required comparisons; the outcomes of comparisons already made, i.e. the available evidence; a spatial plan of to be expected areas, for example: left-hand top corner, left-hand bottom corner, right-hand bottom corner etc.; and, finally, attention-directing features of a configuration, for example: patterns formed by regularities in positions of figures. Future research is needed to examine what type of program is developed by the child and on the basis of which sources of data sampling behavior is controlled by this program when the child grows older. The higher flexibility and efficiency in estimation of older children suggests that the influence of the first three sources may increase with age, while that of the final one may diminish.

Anderson, J.R., & Bower, G.H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123.
Atkinson, J., Campbell, F.W., & Francis, M.R. The magic number 4 ± 0: A new look at visual numerosity judgements. Perception, 1976, 5, 327-334 (a).
Atkinson, J., Francis, M.R., & Campbell. F.W. The dependence of the visual numerosity limit on orientation, colour and grouping in the stimulus. Perception, 1976, 5, 335-342 (b).
Beckwith, M., & Restle, F. Process of enumeration. Psychological Review, 1966, 73, 437-444.
Begg, I. Estimation of word frequency in continuous and discrete tasks. Journal of Experimental Psychology, 1974, 102, 1046-1052.
Begg, I. & Rowe, E.J. Continuous judgments of word frequency and familiarity. Journal of Experimental Psychology, 1972, 95, 48-54.
Bevan, W., & Turner, E.D. Assimilation and contrast in the estimation of number. Journal of Experimental Psychology, 1964, 67, 463-467.
Beyth-Marom, R., & Fischhoff, B. Direct measures of availability and judgements of category frequency. Bulletin of the Psychonomic Society, 1977, 9, 236-238.
Brainerd, Ch.J. Piaget's Theory of Intelligence. Englewood Cliffs, N.J.: Prentice Hall, 1978.
Brooke, J.B. & MacRae, A.W. Error patterns in the judgement and production of numerical proportions. Perception & Psychophysics, 1977, 21, 336-340.
Brown, A.L.& Deloache,J.S., Skills, plans and Self-Regulation. In: R.S. Siegler (Ed.), Children's Thinking: What Develops? Hillsdale, N.J.. Erlbaum, 1978.
Bruner, J.S. Organization of early skilled action. Child Development, 1973, 44, 1-11.
Bryant, P. Perception and understanding in young children. London: Methuen, 1974.
Buckley, P.B. & Gillman, C.B. Comparisons of digits and dot patterns. Journal of Experimental Psychology, 1974, 103, 1131-1136.
Chi, M.T.H. Age differences in memory span. Journal of Experimental Child Isychology, 1977, 23, 266-281.
Chi, M.T.H., & Klahr, D. Span and rate of apprehension in children and adults. Journal of Experimental Child Psychology, 1975, 19, 434-439.
De Finetti, B. The true subjective probability problem. In C.-A.S. STAeL VON HOLSTEIN (Ed.), The concept of probability in psychological experiments. Dordrecht-Holland: Reidel, 1974 (a).
De Finetti, B. The value of studying subjective evaluations of probability. In C.-A.S. STAeL VON HOLSTEIN (Ed.), The concept of probability in psychological experiments. Dordrecht-Holland. Reidel, 1974 (b).
De Zeeuw, G., & Wagenaar, W.A. Are subjective probabilities probabilities? In C.-A.S. STAeL VON HOLSTEIN (Ed.), The concept of probability in psychological experiments. Dordrecht-Holland: Reidel, 1974.
Dixon, P. Numerical comparison processes. Memory and Cognition, 1978, 6, 454-461.
Ekstrand, B.R., Wallace, W.D., & Underwood, B.J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 566-578.
Estes, W.K. The cognitive side of probability learning. Psychological Review, 1976, 83, 37-64.

Fischbein, E. The intuitive sources of probabilistic thinking in children. Dordrecht-Holland: Reidel, 1975.
Flavell, J.H. The developmental psychology of Jean Piaget. Princeton, N.J. Van Nostrand, 1963.
Flavell, J.H. Cognitive Development. Englewood Cliffs, N.J.: Prentice Hall, 1977.
Frith, Ch.D., & Frith, U. The solitaire illusion: an illusion of numerosity. Perception & Psychophysics, 1972, 11, 409-410.
Gelman, R. Conservation acquisition: a problem of learning to attend to relevant attributes. Journal of Experimental Child Psychology, 1969, 2, 167-187.
Gelman, R. The nature and development of early number concepts. In H.W. Reese (Ed.) Advances in Child Development and Behavior. Vol. 7. New York: Academic Press, 1972.
Gelman, R. Counting in the Preschooler: What Does and Does Not Develop? In R.S. Siegler (Ed.) Children's Thinking: What Develops? Hillsdale, N.J.: Erlbaum, 1978.
Gelman, R. & Gallistel, C.R. The child's understanding of number. Cambridge, Massachusetts: University Press, 1978.
Gelman, R., & Tucker, M.F. Further investigations of the young child's conception of number. Child Development, 1975, 46, 167-175.
Ginsburg, H., & Rapoport, A. Children's estimates of proportions. Child Development, 1967, 38, 205-212.
Glanville, A.D., & Dallenbach, K.M. The range of attention. American Journal of Psychology, 1928, 41, 207-236.
Goulet, L.R., & Goodwin, K.S. Development and choice behavior in probabilistic and problem-solving tasks. In H.W. Reese, & L.P. Lipsitt (Eds.) Advances in child development and behavior. Vol. 5. New York: Academic Press, 1970.
Granberg, D. Design as a factor in the study of visual numerousness. Psychonomic Science, 1972, 26, 103-104.
Granberg, D., & Aboud, J.A. Contextual effect in judgements of visual numerousness. American Journal of Psychology, 1969, 82, 221-227.
Guttman, R. Beyond subitizing: is estimation of numerousness a continuous process? Acta Psychologica, 1978, 42, 293-301.
Hasher, L., & Chromiak, W. The processing of frequency information: An automatic mechanism? Journal of Verbal Learning and Verbal Behavior, 1977, 16, 173-184.
Hecox, K.E., & Hegan, J.W. Estimates and estimate-based inferences in young children. Journal of Experimental Child Psychology, 1971, 11, 106-123.
Heymans, P.G. Wikken en Wegen. Unpublished dissertation, Catholic University of Nijmegen, the Netherlands, 1979.
Hintzmann, D.L., & Block, R.A. Repetition and memory: Evidence for a multiple trace hypothesis. Journal of Experimental Psychology, 1971, 88, 297-306.
Howell, W.C. Representation of frequency in memory. Psychological Bulletin, 1973, 80, 44-53.
Huttenlocher, J., & Burke, D. Why does memory span increase with age? Cognitive Psychology, 1976, 8, 1-31.
Inhelder, B., Sinclair, H., & Bovet, M. Learning and the development of cognition. Cambridge, Massachusetts: Harvard University Press, 1974.
Jevons, W.S. The power of numerical discrimination. Nature, 1871, 3, 281-282.
Kahnemann, D., & Tversky, A. Subjective probability: a judgement of representativeness. Cognitive Psychology, 1972, 3, 430-454.
Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. The discrimination of visual number. American Journal of Psychology, 1949, 62, 498-525.
Klahr, D. Quantification processes. In W.G. Chase (Ed.) Visual Information

Processing. New York. Academic Press, 1973.
Klahr, D., & Wallace, J.G. Cognitive Development: An Information Processing View. New York: Wiley, 1976.
Krueger, L.E. Perceived numerosity. Perception & Psychophysics, 1972, 11, 5-9.
La Pointe, K., & O'Donnell, J.P. Number conservation in children below age six: its relationship to age, perceptual dimensions, and language comprehension. Developmental Psychology, 1974, 10, 422-428.
Lawson, G.L., Baron, J., Siegel, L. The role of number and length cues in children's quantitative judgements. Child Development, 1974, 45, 731-736.
Maarse, F.J., Versteegen, W.J.M., Haenen, H.T.M. Gebruiksaanwijzing programma bibliotheek, Psychologisch Laboratorium 2e ed. Rapport 76RG03, Catholic University of Nijmegen, The Netherlands, 1976.
Macnamara, J. A note on Piaget and number. Child Development, 1975, 46, 424-429.
Nash, H. The judgement of linear proportions. American Journal of Psychology, 1964, 77, 480-484.
Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1966.
Nisbett, R.E. & DeCamp Wilson, T. Telling more than we can know: verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.
Peterson, C.R., & Beach, L.R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.
Piaget, J. The child's conception of number. London. Routledge & Kegan Paul, 1952.
Piaget, J. Biology and Knowledge. Edinburg: Edinburg University Press, 1971.
Piaget, J. & Inhelder, B. La genèse de l'idée de hasard chez l'enfant. Paris: Presses Univer. France, 1951.
Pitz, G.F. Response variables in the estimation of relative frequency. Perceptual and Motor Skills, 1965, 21, 867-873.
Potter, M.C., & Levy, E. Spatial enumeration without counting. Child Development, 1968, 39, 265-272.
Pufall, P.B., & Shaw, R.E. Precocious thoughts on number: the long and the short of it. Developmental Psychology, 1972, 7, 62-69.
Pufall, P.B., Shaw, R.E., & Syrdal-Lasky, A. Development of number conservation: an examination of some predictions from Piaget's stage analysis and equilibration model. Child Development, 1973, 44, 21-27.
Schaeffer, B. Skill integration during cognitive development. In A. Kennedy, & A. Wilkes (Eds.) Studies in long term memory. New York: Wiley, 1975.
Schaeffer, B., Eggleston, V.H., & Scott, J.O. Number development in young children. Cognitive Psychology, 1974, 6, 357-379.
Shuford, E.H. Percentage estimation of proportion as a function of element type. exposure type, and task. Journal of Experimental Psychology, 1961, 61, 430-436.
Siegel, L.S. The sequence of development of certain number concepts in preschool children. Developmental Psychology, 1971, 5, 357-361.
Siegel, L.S. The development of concepts of numerical magnitude. Psychonomic Science, 1972, 28, 245-246.
Smither, S.J., Smiley, S.S., & Rees, R. The use of perceptual cues for number judgement by young children. Child Development, 1974, 45, 693-699.
Stevens, S.S., & Galanter, E.H. Ratio scales and category scales for a dozen perceptual continua. Journal of Experimental Psychology, 1957, 54, 377-411.
Svenson, O., & Sjoberg, K. Subitizing and counting processes in young children. Scandinavion Journal of Psychology, 1978, 19, 247-250.
Taves, H.T. Two mechanisms for the perception of visual numerousness. Archives of Psychology, 1941, 37, 1-47.
Trabasso, T., Riley, C.A., & Wilson, E.J. The representation of linear order

and spatial strategies in reasoning: a developmental study. In R.J. Falmagne (Ed.) Reasoning: Representation and Process. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Tversky, A., Kahneman, D. Availability: a heuristic for judging frequency and probability. Cognitive Psychology, 1973, 5, 207-232.

Underwood, B.J. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Vlek, C.A.J. Multiple probability learning associating events with their probabilities of occurrence. In A.F. Sanders (Ed.), Attention and Performance III. Amsterdam: North Holland, 1970.

Vlek, C.A.J., & Wagenaar, W.A. Oordelen en beslissen in onzekerheid. In J.A. Michon, E.G.J. Eijkman, & L.F.W. de Klerk (Eds.) Handboek der Psychonomie. Deventer: Van Loghum Slaterus, 1976.

Vlek, C.A.J., & Werner, H.H. Learning relative frequency distributions: some perceptual and cognitive factors. Journal of Experimental Psychology, 1973, 100, 106-115.

Vos, P.G. Een systematische trend in omissies van vierpuntsfiguren uit de Bourdon-Wiersma test bij kinderen. Nijmegen. vakgroep Ontwikkelingspsychologie, Psychologisch Laboratorium, Katholieke Universiteit, intern rapport 770N02, 1977.

Wagenaar, W.A. Sequential response bias; a study of choice and chance. Unpublished dissertation, University of Leiden, The Netherlands, 1972.

Wagenaar, W.A. Stevens vs. Fechner: a plea for dismissal of the case. Acta Psychologica, 1975, 39, 225-235.

Wallach, L., Wall, A.J., & Anderson, L. Number conservation: the roles of reversibility, addition-subtraction, and misleading perceptual cues. Child Development, 1967, 38, 425-442.

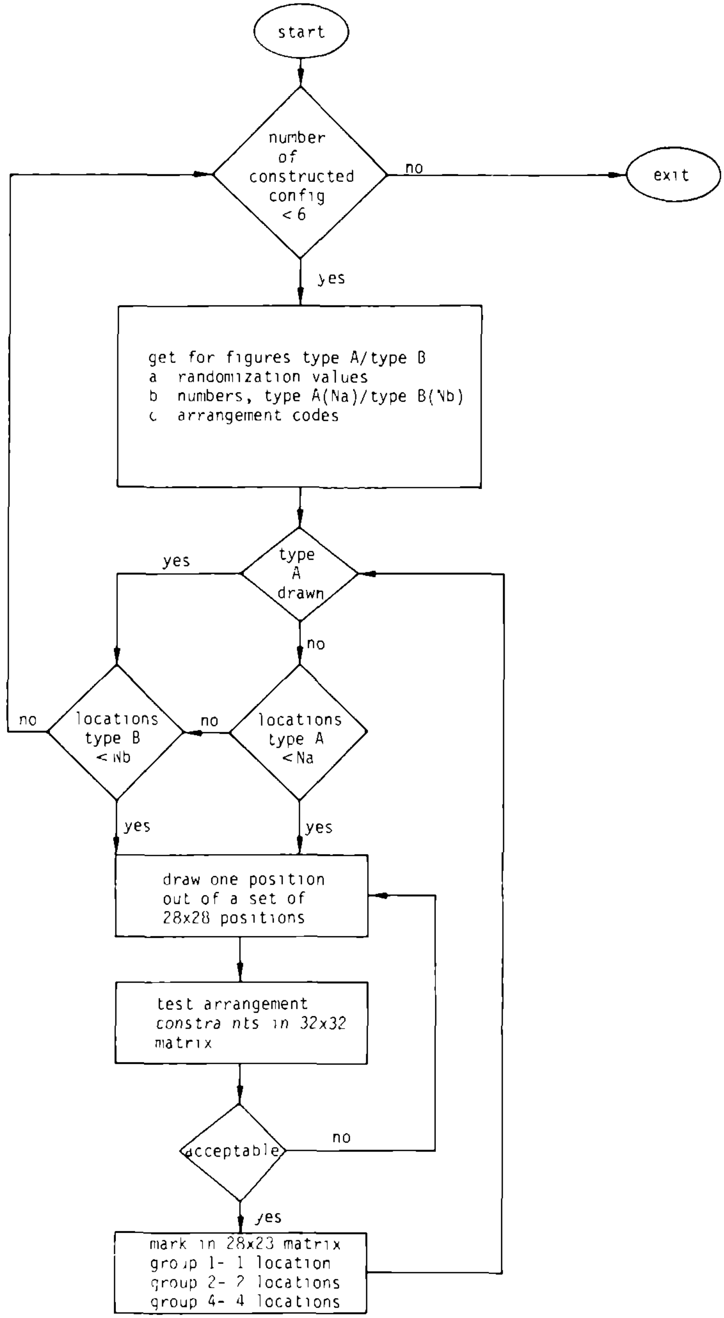
Winer, B.J. Statistical principles in experimental design, 2nd ed. Tokyo: McGraw-Hill, 1971.

Wing, H. The influence of context and stimulus range on the judgement of numerousness. Psychonomic Science, 1971, 24, 199-200.

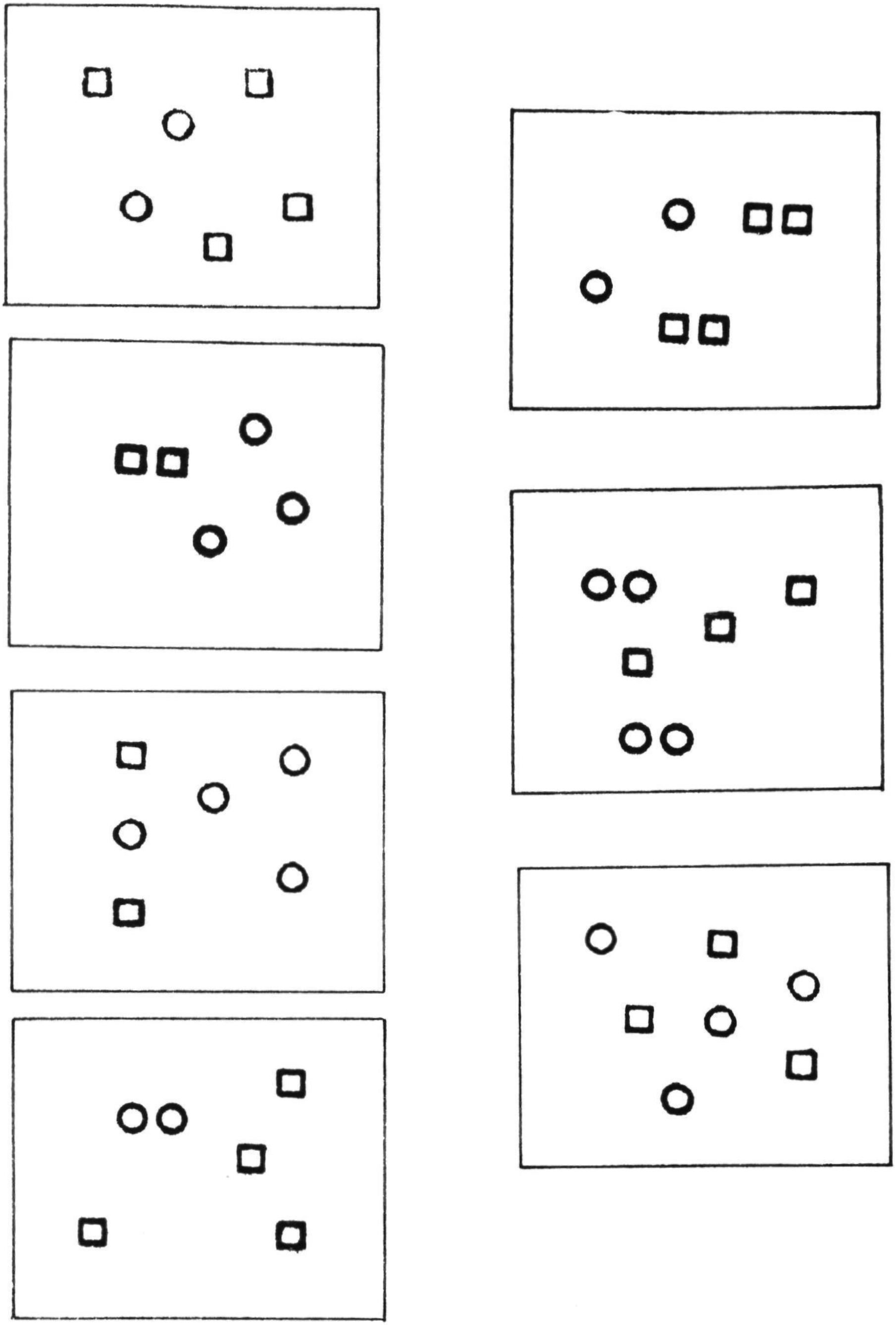
Zimiles, H. A note on Piaget's concept of conservation. Child Development, 1963, 34, 691-695.

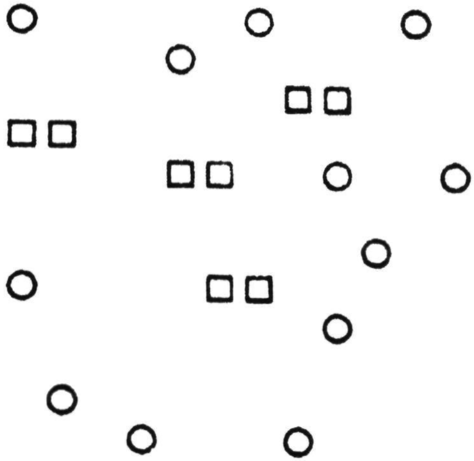
Zimiles, H. The development of association and differentiation of number. Monographs of the Society for Research in Child Development, 1966, 31, 7.

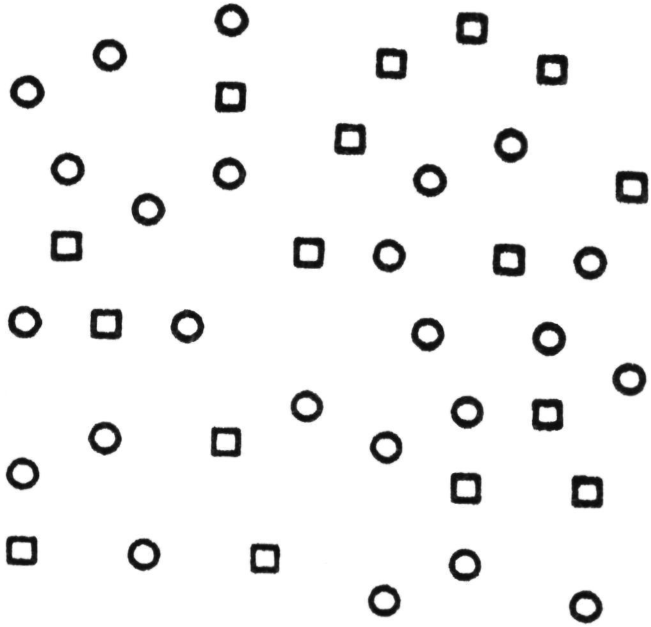
APPENDIX I: flow diagram for construction of stimulus configurations

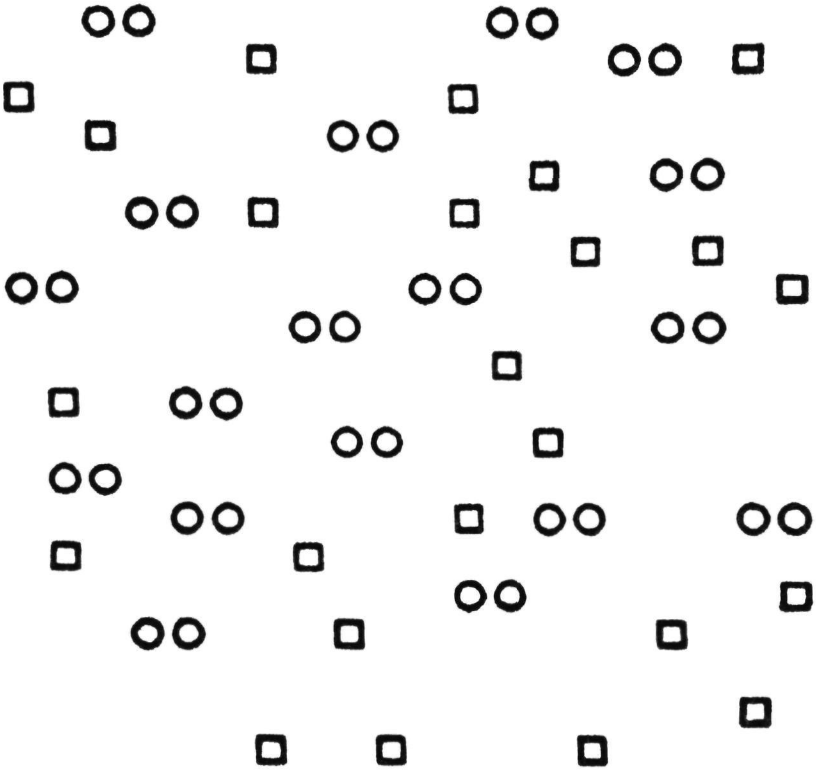


APPENDIX II: stimulus configurations from the instruction series of experiment IV including the given illustrations of samples









The present studies concern the process of estimating discontinuous relative quantities. It was investigated: (1) how the relative quantity of simultaneously presented collections is estimated by persons of different ages, and (2) whether the estimation process does change with age, that is, within the range from about six years to adulthood. The relative quantity judgment examined consisted of quantitative comparison, that is, the establishment of a relation of inequality or equality between collections, or more specifically, the determination of a relation of more, less, or equal.

It was stated that a relative quantity judgment requires at least some representation of the quantities involved. In this respect, number may be considered as very precise and useful, providing the opportunity for exactness. However, the exactness of number cannot be guaranteed when a person has to estimate, for example, because the number of elements is too large and/or the exposure time is too short to count all of them.
The question was raised whether number will be used for representation of quantities when estimation is required, and whether reliance on number is related to age.

The foregoing questions were presented and discussed in Chapter 1. Furthermore, in this chapter the major theoretical perspectives in research on relative quantity estimation were discussed briefly. It was concluded that studies from different perspectives (psychophysics, statistical decision-theory, and verbal learning) generally have shown that, on the average, people are fairly accurate in relative quantity estimation. However, how those quantities are estimated and the conditions that may affect the accuracy of them is not yet clear. Moreover, it was argued that investigation of the estimation process is rather complicated with sequential presentation of items, although in most studies items were in fact presented sequentially. Therefore, in the present studies items were presented simultaneously.

Studies on the estimation of relative quantities in children and adults using simultaneous presentation of items were discussed in Chapter 2. With children, the relevant studies consisted mainly of quantitative comparison studies, and with adults, of proportion estimation studies.

The question was raised whether the development of number conservation suggests new bases for estimating quantities. On the basis of the available evidence it was suggested that children who do not yet show the ability of number conservation will rely on spatial representations of collections. Besides, it was argued that the acquisition of number conservation perhaps indicates that spatial characteristics will be ignored as bases for estimation. In line with this hypothesis it was suspected that from the age of five to seven years a strategy of estimation is developed that is based on numerical rather than spatial information.

In discussing proportion estimation research in adults, the sampling hypothesis was reviewed, and suggestions were made about the quantifiable units of sampling. On the basis of available evidence derived from rather divergent areas of psychological research on quantitative judgment (proportion estimation, numerousness and numerosity judgment and enumeration) it was argued that the quantifiable units of sampling consist probably of small or subitizable groups of objects rather than spatial characteristics or single objects. That is, within a configuration of objects small groups will be identified, each group being of such a size that quantification of its number can be done by subiziting. On inspection of a configuration, numerical information will be acquired by subitizing the number of elements within each identified group. This process was labelled 'sampling by groups'. It was suggested that a configuration is inspected repeatedly depending on its exposure time. It was assumed that on each inspection a number of elements is sampled from each of the relevant categories, and in the case of quantitative comparison a comparison is made between these numbers resulting in some outcome, for example,'more of X'. Outcomes of several comparisons may be combined into an estimation by selecting, for example, the category that occurred as 'more' on most of the comparisons.

Sampling by groups and the development of this process were investigated in four studies. The first study was reported in Chapter 3 and functioned as the basis for the other studies, which were reported in Chapter 4. It was predicted that estimations would be systematically biased, given certain types of arrangements of objects, if sampling by groups was the estimation strategy. That is, given configurations consisting of different types of figures such that subitizable groups may be identified more easily for one than for another category, estimations would systematically favour the category with relatively more easily identifiable groups.

To investigate sampling by groups, configurations were constructed consisting of total number of 120 figures (circles and squares). The configurations varied according to three levels of arrangement or degrees of identifiability of subitizable groups, and in proportion of circles or squares (.40, .50 or .60). Thus, they were constructed according to a 3x3 factorial design. Proportion variation was used to investigate accuracy of proportion discrimination or accuracy of estimation independent of the predicted arrangement effect. Subjects were instructed to judge whether a configuration consisted of more circles or more squares.

In the first experiment, configurations were presented for one, four, or seven seconds each to subjects of four different ages (six, eight, and twelve year olds, and adults). Except for some minor discrepancies, the predicted arrangement effect was found with adults and with twelve year olds but not in eight and six year olds. Moreover, a predicted increase in the accuracy of estimation with presentation time was found in the two older age groups, but not in the two younger ones. Furthermore, as expected, proportions were discriminated at all ages, and accuracy of discrimination or estimation increased with age. Besides, at all ages, latencies increased with presentation time but large differences in latency were found between age groups, especially at the exposure time of seven seconds. Six or eight year olds used less than half the time adults took. In addition, adults used more time for responding than twelve year old children, and estimated relatively more accurately with less bias. However, the better performance of adults compared with that of twelve year olds could be attributed to their relatively longer latency. On the other hand, differences between the two older and the two younger age groups in estimation could not be attributed to the relatively longer latency of the older subjects.

On the basis of these results the general conclusion was drawn that sampling by groups was applied from the age of twelve but relatively more carefully by adults. Some results suggested that the children at the age of eight and six years estimated on a spatial basis, presumably by comparing visual area.

The second study was designed to test the assumption that the arrangement effect would be implied by using sampling by groups. Therefore, sampling by groups was taught to eight year olds who were not expected to show the arrangement effect and to twelve year olds who were expected to apply the teached strategy already. Teaching should induce the arrangement effect at the age of eight but should not affect estimation at the age of twelve.

In order to test the foregoing hypotheses, the stimulus configurations of the first experiment were presented for seven seconds each to a control and training group of children of eight as well as of twelve years old. Before presentation, a short training session was administered to the children in the training groups.

As expected, teaching did not affect estimation at the age of twelve but unexpectedly neither did it at the age of eight. The arrangement effect was observed for the twelve year olds as well as for the eight year olds independent of training. Furthermore, proportion discrimination was found in both age groups independent of training. As expected, latencies of eight year olds appeared to be increased by training, however without any positive effect on accuracy. Children of age twelve used more time for responding than untrained eight year old children, and about the same time as trained eight year olds, but estimated more accurately. Furthermore, twelve year olds seemed to be more adapted to variation in the ambiguity of sample information.

The conclusion was drawn that sampling by groups was applied by the twelve year olds as well as by the eight year olds, but that the children of twelve years old were relatively more able to apply it in an integrated and refined way.

Results of the first two studies suggested that from the age of about eight sampling by groups is used in an increasingly effective way. The third study investigated whether the better performance of older subjects could be attributed to an increased ability of iterative sampling. That is, it was hypothesized that with age subjects are more able to screen a configuration effectively for a relatively longer period of time, inspecting it repeatedly. In line with this hypothesis it was argued that a reduction of the exposure time should minimize age differences. Therefore, in this study the configurations of the previous studies were presented for one second each to children of about eight and twelve years old, and to adults. The one second presentation time was already used in the first study. However, in that study it was suspected that estimations given for the shorter presentation time might have been influenced by prior presentation of the same configurations to the same subject for a longer time.

The hypothesis of iterative sampling was strongly supported by the findings of this study. There were no or small age differences in estimation. Furthermore, the arrangement effect was replicated once more.

The fourth study was intended to induce the arrangement effect in children of about six years old, by teaching sampling by groups to them. Thus, the purpose of this study was comparable to that of the second one. The fourth study showed that the arrangement effect could be induced by teaching sampling by groups to children who did not show it prior to training, and that children of about six are able to learn to estimate in this way.

The results of the four studies were discussed in Chapter 5. The discussion centered around the determinants of estimating on a numberical or non-numerical basis in young children, the question what is developed in estimating, and possible determinants of the development of estimation. Finally, some suggestions were made for future research.

ALGEMENE SAMENVATTING

Deze dissertatie heeft betrekking op het proces volgens welke relatieve hoeveelheden geschat worden. Onderzocht werd:

1) hoe de relatieve grootte van simultaan gepresenteerde verzamelingen voorwerpen geschat wordt door personen van verschillende leeftijd en
2) of het schattingsproces verandert in de periode van zesjarige tot volwassen leeftijd.

Het relatieve kwantiteitsoordeel dat bestudeerd werd was kwantitatief vergelijken. Dit werd gedefinieerd als het bepalen van een relatie van ongelijkheid of gelijkheid, meer specifiek als het vaststellen van een relatie van meer, minder of gelijk.

Er werd gesteld dat relatieve hoeveelheidsoordelen tenminste een of andere representatie van de betrokken kwantiteiten vereisen. In dit opzicht kan aantal gezien worden als zeer precies en bruikbaar. Het geeft de mogelijkheid tot exactheid. Dat aantal exact zal zijn kan evenwel niet gegarandeerd worden wanneer een persoon genoodzaakt is om te schatten, bijvoorbeeld doordat het aantal elementen te groot is en/of doordat de presentatietijd te kort is om alle elementen te tellen.

De vraag werd gesteld of aantal gebruikt zal worden als representatievorm, wanneer het noodzakelijk is om te schatten en of het gebruik van aantal in deze gekoppeld is aan leeftijd.

De hiervoor geformuleerde vragen en de bespreking daarvan werden weergegeven in Hoofdstuk 1. Bovendien werd er in dat hoofdstuk een korte bespreking gegeven van het belangrijkste theoretische perspectief van onderzoek naar het schatten van relatieve hoeveelheden. De konklusie werd getrokken dat onderzoek vanuit verschillende perspektieven (psychophysica, statistische decisietheorie en verbaal leren) in het algemeen aangetoond heeft, dat mensen relatieve hoeveelheden gemiddeld behoorlijk nauwkeurig kunnen schatten. Voorts werd gesteld dat het echter tot nog toe niet duidelijk is hoe die hoeveelheden geschat worden en welke kondities van invloed zijn op de nauwkeurigheid van schatten. Bovendien werd aangevoerd dat sequentiële presentatie van items bestudering van het schattingsproces tamelijk gecompliceerd maakt, ofschoon deze wijze van presentatie in feite in de meeste onderzoeken toegepast werd. In de onderhavige experimenten werden items daarom alle tegelijk aangeboden.

In Hoofdstuk 2 werd onderzoek naar het schatten van relatieve hoeveelheden besproken waarin gebruik gemaakt was van een simultane wijze van presentatie van items. De bespreking was zowel gericht op onderzoek bij kinderen als op dat bij volwassenen. Bij kinderen bestond het relevante onderzoek voornamelijk uit studies van kwantitatief vergelijken en bij volwassenen uit studies van proportie-schatten.

De vraag werd gesteld, of de ontwikkeling van het vermogen van conservatie van aantal bij kinderen mogelijkerwijs betekent dat aantallen op een nieuwe manier geschat gaan worden. Op basis van beschikbare evidentie werd verondersteld dat kinderen die conservatie van aantal nog niet vertonen, zich zullen verlaten op ruimtelijke representaties van verzamelingen. Daarnaast werden er argumenten aangevoerd voor de veronderstelling dat de verwerving van conservatie van aantal wellicht betekent dat spatiële kenmerken niet langer gezien worden als relevante schattingsbases. In overeenstemming hiermee werd het vermoeden geuit dat er vanaf vijf tot zeven jaar een schattingsstrategie ontwikkeld wordt die gebruik maakt van numerieke gegevens in plaats van spatiële.

Bij de behandeling van proportie-schattingsonderzoek bij volwassenen werd de 'sampling-hypothese' besproken. Er werden suggesties gedaan ten aanzien van de kwantificeerbare eenheden bij het trekken van een steekproef. Op basis van beschikbare evidentie werd verondersteld, dat de kwantificeerbare eenheden vermoedelijk niet bestaan uit spatiële kenmerken, of uit afzonderlijke elementen, maar uit groepjes elementen waarbij een groepje zo groot is dat het aantal 'subiteerbaar' is. De argumenten hiervoor werden ontleend aan tamelijk divergente gebieden van psychologisch onderzoek naar kwantitatief oordelen, namelijk het schatten van proporties, het beoordelen van 'numerousness' en 'numerosity' en tellen of opsommen. De veronderstelling hield het volgende in.

Binnen een konfiguratie zullen er groepjes voorwerpen geïdentificeerd worden, waarbij ieder groepje zo groot is dat het aantal elementen onmiddellijk waargenomen kan worden. Bij inspectie van een konfiguratie zal er numerieke informatie verkregen worden door 'subitering' van het aantal elementen van ieder geïdentificeerd groepje. Dit proces werd 'sampling by groups' genoemd. Er werd gesuggereerd dat een konfiguratie, afhankelijk van diens presentatietijd, herhaaldelijk geïnspecteerd wordt. Aangenomen werd dat er bij iedere inspectie een aantal elementen getrokken wordt uit elk der relevante categorieën, waarbij er, in geval van kwantitatief vergelijken, door vergelijking van de verkregen aantallen een uitkomst bepaald wordt, bijvoorbeeld 'meer van X'. De uitkomsten van meerdere vergelijkingen kunnen gecombineerd worden tot een schatting, bijvoorbeeld door de categorie te kiezen die in de meeste vergelijkingen als 'meer' naar voren kwam.

Met een viertal experimenten werd 'sampling by groups' en de ontwikkeling van deze strategie onderzocht. Het eerste experiment is weergegeven in Hoofdstuk 3. Het functioneerde als basis voor de daarop volgende experimenten die weergegeven zijn in Hoofdstuk 4.

Er werd een systematische schattingsafwijking of 'bias' voorspeld bij gebruikmaking van 'sampling by groups', gegeven bepaalde typen ordeningen van voorwerpen. Dit hield het volgende in: bij konfiguraties van figuren van een verschillende categorie, zodanig geordend dat subiteerbare groepjes gemakkelijker geïdentificeerd kunnen worden van de ene dan van de andere categorie, kan verwacht worden dat de categorie met de relatief gemakkelijker te identificeren groepjes systematisch bevoordeeld wordt in een schatting.

Ter bestudering van 'sampling by groups' werden er konfiguraties gekonstrueerd, elk met een totaal van 120 figuren (cirkels en vierkanten). De konfiguraties varieerden volgens drie niveaus van ordening of gradaties van identificeerbaarheid van subiteerbare groepjes. Daarnaast verschilden ze in de proportie cirkels of vierkanten (.40, .50 en .60). De konfiguraties werden dus gekonstrueerd volgens een 3x3 factorieel schema. Proportie-variatie werd toegepast om de nauwkeurigheid van schatten of proportiediscriminatie te kunnen onderzoeken, onafhankelijk van het voorspelde ordeningseffect. Van de proefpersonen werd gevraagd om te beoordelen of er in een konfiguratie meer cirkels of meer vierkanten voorkwamen.

In het eerste experiment werd iedere konfiguratie aangeboden onder drie presentatietijden (1, 4 of 7 sec.) aan proefpersonen uit vier verschillende leeftijdsgroepen (zes-, acht- en twaaljarigen en volwassenen). Het voorspelde ordeningseffect werd - uitgezonderd enkele geringe afwijkingen - gevonden bij volwassenen en bij twaalfjarigen, maar niet bij acht- en zesjarigen. Daarnaast werd een voorspelde toename in nauwkeurigheid van schatten met presentatietijd wel gevonden bij de twee oudste, maar niet bij de twee jongste leeftijdsgroepen. Voorts konden op alle leeftijden de proporties van elkaar onderscheiden worden en nam de nauwkeurigheid van schatten toe met leeftijd. Tevens namen op alle leeftijden de antwoordtijden toe met presentatietijd. Wel verschilden de antwoordtijden aanzienlijk voor de verschillende leeftijdsgroepen, vooral bij de presentatietijd van 7 seconden. Zes- of achtjarigen gebruikten minder dan de helft van de tijd, die de volwassenen namen om te antwoorden. Volwassenen gebruikten ook meer tijd dan twaalfjarigen en schatten relatief nauwkeuriger en met minder 'bias'. De betere prestatie van volwassenen, vergeleken met die van twaalfjarigen, kon echter toegeschreven worden aan hun langere antwoordtijden. Anderzijds konden verschillen in schatten tussen de twee jongste en de twee oudste leeftijdsgroepen niet toegeschreven worden aan de langere antwoordtijden van de oudere proefpersonen.

Op basis van de resultaten van het eerste experiment werd de algemene conclusie getrokken, dat 'sampling by groups' toegepast werd vanaf twaalfjarige leeftijd, maar meer zorgvuldig door volwassenen. Sommige resultaten gaven de indruk, dat door acht- en zesjarigen geschat werd op basis van ruimtelijke kenmerken, vermoedelijk door vergelijking van het visuele oppervlak.

Het tweede onderzoek werd opgezet om de assumptie te toetsen, dat het gebruik van 'sampling by groups' het ordeningseffect zou impliceren. Met het oog daarop werd 'sampling by groups' onderwezen aan achtjarigen, van wie verwacht kon worden dat ze het ordeningseffect niet zouden vertonen. Deze schattingsstrategie werd eveneens onderwezen aan twaalfjarigen van wie verwacht kon worden dat ze 'sampling by groups' reeds toepasten. Het schattingsonderricht zou het ordeningseffect moeten induceren bij achtjarigen. Het zou evenwel geen effect mogen hebben op het schattingsgedrag van twaalfjarigen.

Om de voorafgaande hypotheses te toetsen, werden de stimuluskonfiguraties uit het eerste experiment aangeboden, met een presentatietijd van 7 seconden, aan een trainings- en aan een controlegroep van zowel acht- als twaalfjarige kinderen. Vooraf werd er een korte schattingstraining gegeven aan de kinderen in de trainingsgroepen. Training had - zoals verwacht - geen invloed op het schattingsgedrag van twaalfjarigen, maar, anders dan voorspeld, ook niet op dat van achtjarigen. Op beide leeftijden werd het ordeningseffect gevonden, onafhankelijk van training. Voorts bleek dat de proportiewaarden door de beide leeftijdsgroepen van elkaar onderscheiden werden, onafhankelijk van training. Daarnaast bleken de antwoordtijden van de achtjarigen te zijn toegenomen tengevolge van de gegeven training, overigens zonder een positief effect op de nauwkeurigheid waarmee door hen geschat werd. Tevens gebruikten kinderen van twaalf jaar meer tijd om te antwoorden dan niet getrainde kinderen van acht jaar en ongeveer evenveel tijd als getrainde kinderen van acht jaar. Door de kinderen van twaalf jaar werd echter wel nauwkeuriger geschat. Bovendien leek het schattingsgedrag van twaalfjarigen meer aangepast aan variatie in ambiguïteit van steekproefinformatie.

Als conclusie werd getrokken dat 'sampling by groups' zowel toegepast werd door twaalfjarigen als door achtjarigen. De twaalfjarigen waren echter meer in staat om deze strategie op een relatief geïntegreerde en verfijnde wijze toe te passen.

De resultaten van de eerste twee onderzoeken wezen erop, dat 'sampling by groups' op een toenemend effectieve wijze gebruikt wordt vanaf acht jaar. Met het derde experiment werd nagegaan of de betere prestatie van oudere proefpersonen toegeschreven kon worden aan een grotere vaardigheid in het iteratief trekken van een steekproef. Er werd verondersteld, dat proefpersonen met toenemende leeftijd beter in staat zullen zijn om een konfiguratie voor een relatief lange periode effectief te onderzoeken bij herhaald inspecteren ervan. In overeenstemming met deze hypothese werd gesteld dat een reducering van de presentatietijd leeftijdsverschillen zou moeten minimaliseren. Om het effect van reducering van presentatietijd te onderzoeken werden in Experiment III de konfiguraties uit de vorige experimenten aangeboden aan achten twaalfjarigen en aan volwassenen, onder een presentatietijd van één seconde. Deze relatief korte presentatietijd was reeds toegepast in Experiment I. Er werd toen echter vermoed, dat het schattingsgedrag bij een kortere presentatietijd wel een beïnvloed zou kunnen zijn door eerdere presentatie van dezelfde konfiguraties onder een langere tijd.

De resultaten van het derde experiment ondersteunden de hypothese, dat leeftijdsverschillen in schattingsprestatie toegeschreven kunnen worden aan een grotere vaardigheid van oudere proefpersonen in het iteratief trekken van een steekproef. Er werden geen of alleen zeer kleine leeftijdsverschillen gevonden in schattingsgedrag. Bovendien werd het ordeningseffect eens te meer gerepliceerd.

Met Experiment IV werd nagegaan of het ordeningseffect geïnduceerd kon worden bij zesjarigen door onderricht in de schattingsstrategie 'sampling by groups'. Het doel van dit experiment was dus vergelijkbaar met dat van Experiment II. Het vierde experiment toonde aan, dat het mogelijk is om het ordeningseffect te induceren bij kinderen, die van te voren dit effect niet vertoonden, namelijk, door deze kinderen 'sampling by groups' te onderwijzen. Voorts toonde het onderzoek aan dat zesjarigen in staat zijn om deze schattingsstrategie te leren.

De resultaten van de vier experimenten werden besproken in Hoofdstuk 5. De bespreking centreerde zich rond drie thema's: (1) determinanten van het al of niet schatten op basis van aantal door jonge kinderen, (2) de vraag wat er ontwikkeld wordt bij schatten en (3) determinanten van de ontwikkeling van schatten. Tenslotte werden enkele suggesties gedaan voor toekomstig onderzoek.

CURRICULUM VITAE

Adrianus Willem Smitsman werd geboren op 8 februari 1944 te Zwijndrecht. Na het eindexamen Gymnasium α studeerde hij van 1963-1970 psychologie aan de Vrije Universiteit met als hoofdrichting funktieleer (prof. S.D. Fokkema). Tijdens zijn studie werkte hij op de afdeling funktieleer onder meer mee aan onderwijspsychologisch onderzoek, met name naar het effect van cognitieve structuur-variabelen op het verloop van onderwijsleerprocessen. Dit was tevens het onderwerp van zijn doctoraal scriptie. Na zijn afstuderen in februari 1970 was hij tot juli 1973 werkzaam op het Bureau Sociaal Wetenschappelijk Onderzoek, Ministerie van Defensie/Marine, eerst ter vervulling van zijn dienstplicht, later als wetenschappelijk ambtenaar. In die periode werd onder meer het genoemde onderwijspsychologische onderzoek voortgezet. Sinds juli 1973 is hij als wetenschappelijk medewerker verbonden aan de vakgroep Ontwikkelingspsychologie van de Katholieke Universiteit te Nijmegen.

S T E L L I N G E N

1) Systematische afwijkingen in relatieve hoeveelheidsschattingen kunnen ontstaan doordat de te vergelijken sets verschillen in identificeerbaarheid van kleine groepjes (max. 4-6 elementen).

   Dit proefschrift.

2) De hypothese van Klahr en Wallace dat men bij schatten idiosyncratische representaties van hoeveelheid zou gebruiken is niet houdbaar voor relatieve hoeveelheidsschattingen van tegelijk aangeboden sets. Vanaf ongeveer achtjarige leeftijd blijken dergelijke schattingen op aantal gebaseerd te worden.

   Zie Klahr, D. & Wallace, J.G. Cognitive Development. An Information-Processing View. Hillsdale, N.J.: Erlbaum, 1976, p. 64-66.
   Dit proefschrift.

3) De door jonge kinderen (5-7 jaar) gebruikte strategie waarbij een kwantitatieve vergelijking van tegelijk aangeboden sets gebaseerd wordt op andere kenmerken dan aantal lijkt vooral veroorzaakt te worden door een onvoldoende vaardigheid om aantal te abstraheren uit een patroon en niet zozeer door een ontoereikende conceptualisatie van hoeveelheid.

   Dit proefschrift.

4) Het verschijnsel dat aantallen relatief meer onderschat worden naarmate het aangeboden aantal toeneemt kan gezien worden als een indicatie dat 'sampling by groups' gebruikt wordt als schattingsstrategie.

   Dit proefschrift.

5) De bewering van Elkind en Schoenfeld dat kwantitatieve identiteit een noodzakelijke voorwaarde zou zijn voor kwantitatieve equivalentie is discutabel: ze berust op de aanvechtbare veronderstelling dat transitief redeneren alleen nodig is voor kwantitatieve equivalentie en niet voor kwantitatieve identiteit.

[illegible] Elkind, D. [illegible] [illegible] [illegible]

6) Beschrijvingen van zich ontwikkelend gedrag die niet voldoen aan het door de Zeeuw en Wagenaar beschreven criterium van imiteerbaarheid zijn zinloos.

[illegible] Zeeuw, J. [illegible] Wagenaar, W.A. [illegible] [illegible] [illegible]-holland: [illegible]

7) Het onderzoek van Piaget en Inhelder naar de ontwikkeling van de begrippen toeval en kans toont onvoldoende aan dat deze begrippen zich inderdaad ontwikkelen bij personen. In tegenstelling tot hun theorie, die een overgang van een deterministisch naar een stochastisch paradigma voorspelt, wijzen sommige resultaten van hun onderzoek eerder op een verloop waarbij oudere personen complexere deterministische regels hanteren bij het voorspellen van onzekere gebeurtenissen dan jongere.

[illegible] Piaget, J. [illegible] Inhelder, B. [illegible] Presses [illegible] [illegible] 1951.

8) Het gangbare psychologisch onderzoek naar beslissen onder onzekerheid houdt ten onrechte weinig rekening met de fundamentele behoefte en de mogelijkheid die mensen hebben om hun omgeving te beïnvloeden.

9) Bestudering van ontwikkelingsverschijnselen in de eerste of tweede levenshelft van personen kan de facto pas achteraf plaatsvinden.

10) Het verdient aanbeveling om in het diagnostiek onderwijs studenten alert te maken voor het oordeelsproces van de diagnost zelf.

11) Het feit dat schepen alleen be-mand kunnen worden zegt niets over de zeewaardigheid van vrouwen.

Nijmegen, 8 februari 1980 A.W. Smitsman